THREE ESSAYS ON INEQUALITIES IN INCOME AND HEALTH

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Jeffrey Harold Larrimore

May 2010

THREE ESSAYS ON INEQUALITIES IN INCOME AND HEALTH

Jeffrey Harold Larrimore, Ph. D.

Cornell University 2010

This dissertation considers several aspects of the distribution of income and income inequality. It does so by improving estimates of inequality between demographic groups, analyzing factors contributing to US income inequality trends, and estimating the impact of income on health outcomes for individuals in the lower tail of the income distribution.

Most empirical studies of earnings and income inequality across demographic groups are based on data from the public use March CPS. However, censoring of high incomes in this data prevent researchers from observing the full distribution. The first essay uses internal CPS data to illustrate how topcoding results in the understatement of income and earnings gaps between men and women, Blacks and Whites, and people with and without disabilities. It also demonstrates how a new series of mean incomes for topcoded observations can be used in conjunction with public use CPS data to closely approximate these internal results.

The second essay considers the factors accounting for trends in household income inequality. Using a shift-share approach, this essay analyzes whether income inequality shifts are accounted for by male and female earnings distribution changes or by changing household characteristics. It illustrates that the factors contributing to the rapid rise in household income inequality in the 1970s and 1980s differ substantially from those contributing to slower increases in the 1990s. In contrast to findings for the 1970s and 1980s, in more recent years increases in male earnings inequality largely

account for household income inequality trends while declines in the correlation of spouses' earnings have mitigated household income inequality growth.

The final essay shifts from considering income inequality to the impact that income has on the health of low income individuals. Health economists have long observed a positive relationship between health and income but the reason for this relationship is unclear. Using exogenous variation in income from state-level differences in the Earned Income Tax Credit, it observes the impact on morbidity of an exogenous increase in income for low income individuals. The results find only weak evidence that the increases in income result in improvements in self-reported health status or the prevalence of functional limitations.

BIOGRAPHICAL SKETCH

Jeffrey Larrimore was born in Bryn Mawr, Pennsylvania in 1982. He attended Davidson College as an undergraduate, where he received his B.A. in May of 2004 with a double major in Economics and Political Science. After a year working as an Analyst for Accenture, he continued his education at Cornell University beginning in August 2005. He received his M.A. in Economics in May of 2009 and his Ph.D. in Economics in May of 2010. He has accepted a position as an Economist with the US Congress Joint Committee on Taxation.

To my wife, Laura Larrimore, whose patience, understanding, and support over the past five years made this endeavor possible; and to my parents, Pat and Dale Larrimore, my sister, Lisa Larrimore Ouellette, and my Grandmother, Kathleen Ann Hebeler, who always pushed me to succeed and to strive for something more.

ACKNOWLEDGMENTS

There are many people to whom I owe thanks for their generous help and support with this dissertation. First, I want to thank my Committee Chair, Richard Burkhauser, for his invaluable mentorship and guidance over the past four years. He first introduced me to inequality research and his guidance as I pursued this field of research allowed me to grow as a scholar much more rapidly than would have otherwise been possible. For this I am profoundly grateful.

I am also extremely appreciative of the support provided by John Cawley and Donald Kenkel. Their advice both on my dissertation committee and in the classroom has helped me to develop my research interests and begin pursuing those interests as I embark on my career as an economist. Both this dissertation and my abilities as a researcher have benefitted greatly from their guidance.

While often unheralded, I also want to express my appreciation for my classmates and colleagues in the Department of Economics and Department of Policy Analysis and Management at Cornell University. This work has benefitted from many of their comments, and especially those of Joseph LeCates who has been remarkably selfless in his willingness to spend his time offering feedback on my research.

Finally, I gratefully acknowledge Cornell University's Department of Policy Analysis and Management and the National Institute for Disability and Rehabilitation Research (award nos. H133B040013 and H133B031111) for their financial support as I was writing this dissertation. I also appreciate Arnie Reznek, Laura Zayatz, Ian Schmutte, the Cornell Census RDC Administrators, and all their U.S. Census Bureau colleagues for their help accessing and using the restricted access Census data for the first two chapters of this dissertation. Any opinions and conclusions expressed herein are those of the author and do not necessarily represent the views of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

PREFACE

Economists and policymakers often use trends in inequality as a measure of social policy success. But such trend outcomes can be sensitive to both the concept of inequality used and the way the concept is implemented. Inequality can be measured across segments of the population, such as the labor earnings gap between males and females or Blacks and Whites, or on a societal level irrespective of individual characteristics. Inequality can also be measured on different types of income, such as labor earnings inequality, which capture inequalities in how individuals are compensated for their time in the labor market or household income inequality, which better captures inequalities in individuals' resources available for consumption. Alternatively, inequality may exist in aspects of society that are separate from income – such as the inequality of the health of individuals. Each essay of this dissertation examines the levels and trends in various types of inequality using CPS data that corrects for differences in the method used to collect the data over the past 40 years, or explores the factors contributing to these inequalities.

The first essay considers the measurement of inequality in income and labor earnings across demographic segments of the population and demonstrates the sensitivity of these results to data reporting procedures used by the Census Bureau. Since most empirical studies of trends in income and earnings inequality across subgroups of the population are based on data from the public use March Current Population Survey (CPS), it is important for this data to accurately reflect the full income distribution in the United States. However, censoring of high-incomes in the public use March CPS prevents researchers from observing the full income distribution.

Using the internal CPS, this essay reexamines estimates of labor earnings gaps between men and women, Blacks and Whites, and individuals with and without disabilities. It illustrates how previous estimates using the public use March CPS data will both understate the levels of inequality and misstate the trends because of the Census Bureau's topcoding procedures. It also shows how population-level estimates of inequality and measures of inequality within demographic groups will similarly be affected by topcoding in the public use data.

The problems with the CPS identified in this essay, however, do not mean that researchers should abandon the public-use March CPS as a tool for analyzing inequality in the United States but rather that it is essential to utilize new tools for capturing top incomes when using this topcoded data. This essay describes how a new series of mean incomes for censored observations based off of the internal CPS data can be used in conjunction with the public use CPS data to allow researchers to closely approximate the earnings and income gaps found in the internal CPS data. Thus, by using this series of mean incomes researchers can largely mitigate the topcoding problems which led to the underestimates of earnings and income gaps described here.

Having established that previous estimates of income inequality may be understating inequality and misstating the trends, the second essay uses improved household income inequality trends based off of the internal March CPS and turns to the question of what factors account for these trends. While much of the research examining the rise in inequality in the United States since the 1960s has focused on labor earnings inequality, there is little evidence regarding how closely these labor earnings inequality trends correlate to broader measures of income inequality. This essay first compares male and female labor earnings inequality to that of household income. It then uses a shift-share analysis to analyze the change in income inequality accounted for by changes in male and female earnings distributions and changing household characteristics.

When using the shift-share analysis, it is evident that the factors contributing to the rapid rise in household income inequality in the 1970s and 1980s differ substantially from those contributing to the slower increase in the 1990s. In contrast to findings for the 1970s and 1980s, in more recent years increases in male earnings inequality largely account for the changes in household income inequality. This is the case since changes in the correlation between spouses' earnings have mitigated household income inequality growth. Hence, researchers focusing only on male labor earnings inequality as a factor important to household income inequality will understate the rise in income inequality in the United States in the 1980s but overstate the growth in income inequality in the 1990s. Since household income inequality more closely captures inequality in the resources available for individuals to consume goods and services than do labor earnings, these findings are important for understanding the more complete set of mechanisms which may influence inequality in access to material goods and services in the United States.

The final essay of this dissertation shifts from considering inequality in income to inequality of health status. Health economists have long observed that individuals with low incomes tend to be in worse health than individuals with high incomes. As a result, when considering inequality across multiple dimensions this positive relationship may suggest a more unequal society given that individuals with more income not only have greater financial resources but also have better health. This essay first documents the existence of this positive relationship in two large, nationally representative samples – the March CPS and the Survey of Income and Program Participation. It then demonstrates that the relationship is stronger among individuals at the bottom tail of the income distribution.

However, in order to understand the implications of this relationship it is also valuable to explore the mechanisms behind it. It is unclear whether this positive relationship exists because increased income allows individuals to purchase more health inputs contributing to their better health, or because healthy individuals are more productive and can obtain higher wages in the labor market, or because a third factor contributes to increases in both health and income. This essay considers the first of these three mechanisms. Using the exogenous variation in income that results from state-level differences in the Earned Income Tax Credit, it observes the impact on health status of an exogenous increase in income for low income individuals. The results find only limited evidence that increases in income reduce morbidity for low income individuals.

CHAPTER 1

USING INTERNAL CPS DATA TO REEVALUATE TRENDS IN INCOME INEQUALITY AND EARNINGS GAPS

Abstract

Most empirical studies of trends in earnings inequality across subgroups of the population are based on data from the public use March Current Population Survey (CPS). However, censoring of high-incomes in the public use March CPS prevents researchers from observing the full income distribution. Using the internal CPS, this chapter shows that inconsistent topcoding in the public use data will understate labor earnings and household income gaps between men and women, Blacks and Whites, and between individuals with and without disabilities. It also illustrates how topcoding has reduced perceptions of inequality across the entire population, along with inequality among individuals with given sets of demographic characteristics. Finally, using a new series of mean incomes for censored observations based off of the internal CPS data, it demonstrates how researchers with access to only the public use data can obtain inequality estimates that more closely approximate the earnings and income inequality found in the internal CPS data.

1.1 Introduction

The March Current Population Survey (CPS) public use files are one of the primary data sources used to study income inequality trends in the United States. Researchers using these data files have extensively considered measures of labor earnings inequality and household income inequality across demographic groups, such as the female-male or black-white earnings gap; within demographic groups, such as the earnings inequality of just working age men; and for the population as a whole. Some key findings of this research include the rapid decline in the male-female labor earnings gaps in the 1980s (see, e.g. Blau and Kahn 1997, Blau and Kahn 2000, and Card and DiNardo 2002), the lack of decline in the Black-White labor earnings gap from the mid 1970s to the early 1990s (Juhn, Murphy, and Pierce 1991, Couch and Daly 2002, and Juhn 2003) and the rise in household income inequality over the same period (Gottschalk and Danziger 2005, Daly and Valetta 2006, and Burkhauser et al. Forthcoming).

While the Current Population Survey is a valuable resource for such studies given its large nationally representative sample of households and its extensive battery of income questions, it suffers from an important limitation. To protect the confidentiality of its respondents the Census Bureau topcodes the highest values from each source of income that it collects when it reports the income in the public use CPS data. This procedure greatly limits the information available to researchers about the incomes at the upper tail of the income distribution and may distort observed income inequality trends.

With access to more complete data on the full distribution of income in the United States, including information on incomes above the public use topcode thresholds, this paper reconsiders US inequality trends and explores the impact that Census Bureau topcoding procedures have had on previous estimates. It first illustrates that level and trends of labor earnings and household income gaps between men and women, Blacks and Whites, and people with and without disabilities may be misstated as a result of topcoding. It then demonstrates that the level and trends for population-wide and household income inequality are also sensitive to topcoding. Finally, it decomposes population-wide household income inequality measures into within- and between-group inequality to determine the extent to which each of these components are sensitive to topcoding.

1.2Measuring income and inequality

Measuring Income and Earnings. There are two distinct income concepts studied by inequality researchers which must be defined precisely since they will be referred to regularly in this paper and their terminology is often confused. The first is labor earnings (also referred to as earnings), which includes all money received by an individual for his or her labor market activities over the course of a year. The vast majority of labor earnings come from wages and salaries, but it also includes self-employment and farm earnings.[1]

When evaluating measures of labor earnings inequality, the sample is restricted to working age individuals (age 22-62) to limit the influence of schooling and retirement decisions on results. Additionally, to further limit the impact of work-intensity decisions, only individuals working full-time and full-year are included.[2] Thus, references to labor earnings inequality in this paper reflects inequality among individuals who are participating full time in the labor market rather than among the entire population. Counts of the number of people in each population group analyzed, and the number of people in each group who are of working age and work full-time are provided in Appendix Table 1.1.

The second income concept is household income which includes all money received from any individual in the household and from any income source over the course of a year. As it includes both income from outside the labor market and income

---

[1] Since self-employment earnings reflect the value of an individual's labor market activities, their inclusion better reflects the difference in how an individual's labor market activities are valued. Devine (1994) demonstrates that excluding self-employment and farm earnings leads to an understatement of the gender labor earnings gap since the gap between male and female earnings is larger among self-employed workers.

[2] While this limits the impact of work-intensity decisions on results, it does not completely alleviate the problem if there are selection effects on the type of people who chose to work full-time. Including all working age individuals or including all individuals working for pay regardless of intensity would likely increase the gender earnings gap since women work full-time at lower rates than men. Similar results should occur for earnings gaps based on race or disability status. Focusing on hourly or weekly wages, rather than annual wages, would have mixed effects depending on how the earnings gap for part-time workers compares to that of full-time workers.

from other household members, household income is a substantially broader definition of income than labor earnings. When considering household income inequality, this paper follows the standard procedure in the literature of using the individual as the unit of analysis and examining their size-adjusted household income. In examining household income at the person level, it is assumed that total household income is shared equally among household members regardless of its source.[3] Since income is assumed to be shared equally among all household members, unlike with labor earnings the sample is not restricted by age and the results reflect inequality for the entire population.[4]

Measuring Inequality. Two methods are used for measuring inequality in earnings and income. The first is the mean income or earnings gap between population subgroups, calculated as one minus the ratio of the mean income of the two groups. Given the intuitive nature of this metric, income and earnings gaps and the related earnings ratios are two of the most common metrics for evaluating inequality between population groups. For example, earnings ratios are reported by the Census Bureau as an illustration of inequality in labor earnings by gender in their annual report on income in the United States (Denavas-Walt, Proctor, and Smith 2009). They also are used by researchers interested in inequality in labor earnings by gender (see e.g. Blau and Kahn 2000)

To limit the impact of outliers, the Census Bureau and others often use median rather than mean income when reporting the income gaps. However, using median

[3] Following the standard convention in the household income inequality literature, size-adjusted household income is calculated as total household income is divided by the square root of the number of individuals living in the household. This approach accounts for household size and sharing of resources within the household while also accounting for economies of scale in household consumption. See e.g. Atkinson, Rainwater, and Smeeding (1995) and Burkhauser, Feng, and Jenkins (2009).

[4] While individuals of all ages are included, those living in group quarters or in the military are excluded. This matches the sample used by Burkhauser, Feng, and Jenkins (2009) and Burkhauser et al. (Forthcoming).

income comes at the cost of focusing only on the midpoint of the distribution. As a result, if substantial shifts in the relative distribution of groups occurred at either tail of the distribution, a comparison of the median will miss these changes. Additionally, since income and earnings distributions are positively skewed in all years, analyses of the means give relatively more weight than the median to changes in the upper tail of the distribution. So for researchers interested in this portion of the distribution, the mean is better able to capture differences between groups and changes over time. Since this paper focuses on the upper tail of the distribution where most topcoding occurs, it uses mean income and earnings, which better reflects changes throughout the entire distribution better captures the impact of topcoding on these gaps.

Population-wide inequality and the level of within-group and between-group inequality are measured using the *GE(1),* or Theil index, which is a middle-sensitive member of the Generalized Engropy class of inequality indices. Like the Gini coefficient, Generalized Entropy inequality indices satisfy the four desirable properties of inequality indices: scale invariance, replication invariance, symmetry (or anonymity), and satisfying the Pigou-Dalton transfer principle (Jenkins and Van Kerm 2009). However, unlike the Gini coefficient, Generalized Entropy inequality indices are additively decomposable into a within-group component reflecting the income inequality among members of a demographic group and a between-group component reflecting the income inequality across two or more demographic groups. As a result, Generalized Entropy inequality indices such as the Theil are more suitable than the Gini for analyzing the extent to which changes in population-wide inequality has been attributable to inequality within demographic groups of the population and to what extent it is attributable to inequality shifts between demographic groups.[5]

[5] Similar decompositions to that performed for the Theil are provided in Appendix Tables 1.4 and 1.5 for the other two most common Generalized Entropy indices: GE(0), the Mean Log Deviation, and GE(2), half of the squared coefficient of variation.

The within-subgroup component, $GE_w(\alpha)$, of each Generalized Entropy inequality index is calculated as:

$$GE_w(\alpha) = \sum_{k=1}^{K}[(v_k)^{1-\alpha}(s_k)^{\alpha}GE_k(\alpha)] \qquad (1.1)$$

where v_k is the subgroup's share of the total population, $(\frac{N_k}{N})$, s_k is the subgroup's share of the total income, $(\frac{Y_k}{Y})$, and $GE_k(\alpha)$ is the Generalized Entropy inequality measure for the subgroup treating it as an independent population. For the Theil index, where $\alpha = 1$, the subgroup's share of the total population drops out and the within-subgroup component of inequality is:

$$GE_w(1) = \sum_{k=1}^{K}[(s_k)GE_k(1)] \qquad (1.2)$$

To calculate the between-subgroup component, $GE_B(\alpha)$, each member of a subgroup is assumed to have identical incomes equal to the mean income of the group. The between-group inequality is then the population's generalized entropy value if this assumption were true. Thus, for the between-group inequality there is no remaining inequality of incomes within the groups and all inequality is resulting from differences in the between-group mean incomes. The total inequality in the entire population is simply the sum of these within- and between-subgroup components:

$$GE(\alpha) = GE_w(\alpha) + GE_w(\alpha) \qquad (1.3)$$

For further details on the Generalized Entropy inequality measures and its decomposition, see Cowell (2000) and Jenkins and Van Kerm (2009).

1.3 Topcoding in the March CPS

The analyses in this paper are based on data from both the public use and internal March CPS data files. These are the primary data sources used for analyzing inequality in the United States. In the public use CPS, each source of income—11 prior to 1987, and 24 since then—has a topcode.[6] When the amount of income reported for an income source in the internal CPS is below or equal to the topcode threshold, the exact amount of income is recorded for that income source in the public use CPS. But when the amount is greater than the topcode threshold, the reported income from that source is suppressed and is recorded as the value of the threshold.

Topcoded income sources include not just labor earnings but also non-labor income including interest, dividends, supplemental security income, workers' compensation, and all other income recorded on the March CPS survey. Because topcodes on non-labor income can be quite low, topcoding affects not only those with very high incomes but also those with relatively modest total income whose income from one or more sources exceeds the topcode threshold (See Appendix Tables 1.2 and 1.3 for a full list of the topcode thresholds over time). In addition to topcoding each income source in the March CPS, the Census Bureau topcodes income reported in CPS surveys from other months, such as the usual weekly earnings reported in the surveys filled out by outgoing rotation groups. This further topcoding prevents researchers from obtaining additional income information from other questions in the CPS.[7]

When considering income inequality for the entire population, it is easy to see how topcoding reduces observed levels of income inequality. Since topcodes reduce

[6] Each CPS survey captures income from the previous year. In this paper all references are to the income year. Thus, a reference to 1987, for example, refers to income received in 1987 and recorded in the 1988 March Current Population Survey.

[7] While much of the earlier research on topcoding focuses on CPS data, issues of confidentiality make topcoding of income a necessity for most national datasets both in the US and internationally. For instance the National Longitudinal Survey of Youth (NLSY97) topcodes some of its income sources as does the Panel Study of Income Dynamics (PSID), the Survey of Income and Program Participation (SIPP), and the American Community Survey (ACS). In Great Britain, in order to comply with the 2007 Statistics and Registration Services Act, the Annual Population Survey and the Quarterly Labour Force Survey have introduced topcodes on earnings data in their main public release files. In numerous countries including Germany, Austria, and the United States, the wage data that are available from social insurance or social security administrative registers are censored at the earnings level corresponding to the upper limit to social insurance contributions.

the level of observed income at the top of the distribution, the dispersion of income in the population is artificially compressed. As a result, the presence of topcoding will necessarily reduce measures of inequality for the entire population. The more severely the income distribution is topcoded, the more the observed level of inequality will be reduced.

Less obvious are the impacts that topcoding has on cross-group measures of inequality such as the ratio of mean earnings between men and women. If the income distribution of men is identical to that of women then individuals in both groups will be topcoded at the same rate. In this case while topcoding will reduce the mean earnings of both groups, it is reduced by the same amounts for both men and women and thus would have no effect on mean earnings gaps. However, if women are topcoded at lower rates than men then topcoding will artificially raise the ratio of their mean income relative to men's because women's observed mean income will be less artificially depressed. Thus, topcoding can also influence levels of cross-group inequality depending on how topcoding rates differ across the groups.

In addition to influencing the levels of inequality in a given year for the reasons described above, topcoding will also distort trends in inequality because topcode thresholds vary inconsistently over time. Rather than increase topcodes with inflation or real-earnings growth to keep a constant percent of the population topcoded in all years, the Census Bureau keeps the topcode threshold constant in nominal terms in most years with periodic substantial increases in the topcode thresholds. This has resulted in marked variations in topcoding rates over time. In years where the fraction of the population topcoded increases, the distortionary effects of topcoding increase and there is an artificial reduction in the observed inequality in the population. Conversely, when the threshold increases and the fraction of the population topcoded declines and less income is suppressed, there is an artificial increase in observed population-wide inequality measures. See Levy and Murnane (1992) for a more formal statement of this problem.

Recognizing these problems from topcoding, in 1995 the Census Bureau offered a partial solution for the underreporting of top incomes at the top of the distribution. Prior to this time, the Census Bureau simply replaced the value for each source of an individual's income that was topcoded with the topcode threshold. Since 1995, the Census Bureau began replacing topcoded incomes in the Public Use CPS with its cell mean value—the mean value of all topcoded values from that source of income.[8] Since cell means were not provided retroactively in years prior to 1995, using the public use CPS data without taking this major change in reported income values into account results in a sizable increase in 1996 and beyond in their measured income due to more accurate reporting of their incomes since then. Hence, while the use of cell means after 1995 causes the public use CPS to conform better to the internal CPS, not taking this improvement in measurement into account will grossly overestimate actual increases in income after 1995 compared to prior years.

1.4 Prevalence of topcoding by gender, race, and disability Status

Topcoding on labor earnings sources: Table 1.1 shows the percentage of full-time, full-year workers in each population group who have labor earnings topcoded in the public use CPS in the trough years of each business cycle since 1975.[9] Although it is not a trough year, 1992 is included in Table 1.1 and all subsequent tables describing

[8] For individuals topcoded on one of the sources of labor-earnings, the Census Bureau provided further detail on top incomes by providing the mean of topcoded incomes of individuals topcoded on the same source of income with the same race, gender, and employment status. See Larrimore et al. (2008) for details.

[9] The starting and ending years of a business cycle are somewhat arbitrary. Rather than define them directly by changes in macroeconomic growth, they are defined here using troughs in income which will, in general, lag macroeconomic growth. The findings in this paper are not sensitive to reasonable changes in business cycle trough years.

Table 1.1: Percentage of Full-Time and Full-Year Workers of working-age who are Topcoded on their Labor Earnings by Gender, Race, and Disability Status

Income Year	Female (1)	Male (2)	Ratio (1)/(2)	Black (3)	White (4)	Ratio (3)/(4)	Disabled[1] (7)	Not Disabled[1] (8)	Ratio (7)/(8)
1975	0.02%	1.22%	0.01	0.00%	0.95%	0.00			
1982	0.18%	1.96%	0.09	0.34%	1.45%	0.23	1.50%	1.30%	1.15
1992	0.42%	3.12%	0.14	0.34%	2.31%	0.15	0.90%	2.03%	0.44
1993	0.66%	3.71%	0.18	0.61%	2.87%	0.21	0.70%	2.51%	0.28
2004	0.59%	2.15%	0.27	0.49%	1.80%	0.27	1.64%	1.50%	1.09
2007	0.93%	2.46%	0.38	0.83%	2.20%	0.38	0.77%	1.82%	0.42

Source: Author's calculations using public use March CPS data.
[1]Disability status is not available prior to 1980.

results for trough years of business cycles. As is discussed in more detail below, Census Bureau data collection procedures were redesigned after 1992 which reduced data comparability across these two years. So 1992 is the last year of the earlier procedures and 1993 represents both the trough year of the 1990s business cycle and the first year of the new procedures. Table 1.1 also includes 2007 which is not a trough year but is the last year available in the internal CPS data.

As can be seen in Table 1.1, while the level of topcoding varies by race and gender; its prevalence has increased greatly over the past 35 years for individuals of each group. In 1975, topcoding on labor earnings was negligible for female or black workers. However, by 2007 close to 1 percent of women (column 1) and Blacks (column 4) were topcoded on labor earnings and even higher percentages of men (column 2) and Whites (column 5) were topcoded.

In contrast, the rate of topcoding on labor-earnings for full-time, full-year workers with disabilities has actually fallen since the 1982 trough year (column 7).[10]

[10] Disability status is measured in the March CPS data starting in 1980 using the work limitation question: "Does anyone in this household have a health problem or disability which prevents them from working or which limits the kind or amount of work they can do", with a follow-up question to determine the disabled member of the household if the response is affirmative. While this work limitation-based measure of disability is not as nuanced as those found in other data sets, it is widely used in the economics literature to capture the working age population with disabilities (see, e.g., Acemoglu and Angrist, 2001 and Hotchkiss, 2004). Since this question asking about work limitations

However, this is partially due to the selection effects of which individuals with disabilities are willing and able to work full-time since labor earnings are evaluated here for full-time, full-year workers only. The fraction of all working age individuals reporting a disability who work full-time fell from 13.06 percent in 1982 to 8.35 percent in 2007. As a result of this decline, the types of individuals with disabilities in the workforce have likely changed thus impacting the topcoding rates. Additionally, since the sample of individuals with disabilities who work full-time is relatively small, the observed fluctuations in topcoding rates result from only mild shifts in the number of individuals with disabilities reporting incomes above the topcode thresholds.[11]

In addition to the general increases in labor earnings topcoding rates over time for each demographic group, in any given year there are noticeable differences in topcoding rates between these groups. This can be seen in Columns 3, 6, and 9 of Table 1.1, which illustrates the ratio of topcoding rates across groups. In 2007, women were topcoded 33 percent as much as men, up from only 2 percent as much in 1975. In 2007 Blacks were topcoded 37 percent as much as Whites, compared with 1975 when no Blacks were topcoded. Given these large differences in topcoding rates across groups, one should expect that topcoding will impact levels of earnings gaps. Additionally, the changes in relative topcoding rates over time should factor into the trends in inequality with earnings gaps being suppressed to a greater degree in years when there is more topcoding overall and in years when there is a greater disparity in relative topcoding rates.

Topcoding on any source of household income: As previously described, labor

---

was not asked prior to 1980 or of individuals under the age of 15, results comparing individuals with and without disabilities are not available prior to that time or for children under the age of 15.

[11] For example, the shift from 2004 to 2007 in the fraction of working age full time workers with a disability was a shift from 13 surveyed individuals with topcoded income to 7 individuals. In comparison, in 2007 there were 253 women working full time topcoded in the March CPS on any labor earnings and 62 Blacks topcoded on any labor earnings.

Table 1.2: Percentage of individuals who are Topcoded on any source of household income by Gender, Race, and Disability Status

Income Year	Female (1)	Male (2)	Ratio (1)/(2)	Black (3)	White (4)	Ratio (3)/(4)	Disabled (7)	Not Disabled (8)	Ratio (7)/(8)
1975	0.84%	0.97%	0.87	0.01%	1.07%	0.01			
1982	1.34%	1.52%	0.88	0.28%	1.65%	0.17	0.54%	1.51%	0.35
1992	2.20%	2.55%	0.86	0.35%	2.86%	0.12	0.54%	2.52%	0.21
1993	2.73%	3.13%	0.87	0.92%	3.52%	0.26	0.78%	3.05%	0.26
2004	4.56%	4.90%	0.93	2.10%	5.79%	0.36	3.34%	5.05%	0.66
2007	5.73%	5.94%	0.97	2.73%	7.37%	0.37	4.24%	6.27%	0.68

Source: Author's calculations using public use March CPS data.
[1]Disability status is not available prior to 1980. Comparisons of household income for people with and without a disability excludes individuals age 15 and under, for whom disability status is not captured in the March CPS

earnings are just one component of the total income available to an individual. Thus, while labor earnings inequality is important in understanding how individuals are compensated differently for their time, it is important to consider broader measures of inequality as well. For this reason, in the income inequality literature it is commonly assumed that income is shared equally across members of a household when calculating income inequality statistics (see e.g. Karoly and Burtless 1995, Atkinson, Rainwater, and Smeeding 1995, and Burkhauser, Feng, and Jenkins 2009). Table 1.2 therefore considers the prevalence of topcoding on any source of household income over time for each of the demographic groups discussed above. Since sharing of resources is assumed, Table 1.2 includes all individuals rather than just full-time, full-year workers of working age.[12]

There is a noticeable difference between the relative gender topcoding rates on household income and those seen for labor earnings. While working females are topcoded on their own labor earnings at less than half the rate of working men in all years, for household income female topcoding rates are always at least 85 percent of

[12] For comparisons of individuals with and without disabilities, only individuals aged 15 and older are included because no information on one's disability status is available for individuals under age 15.

that of men's (column 3). In contrast, the difference in topcoding rates across race and disability status is similar or larger for household income than for labor earnings of working individuals.

There are several potential reasons that could lead to the relationship between labor earnings and household income topcoding rates by gender being different from the relationship seen by race or disability status. These explanations include relative topcoding rates being affected by the expansion of income sources to include all income rather than just labor earnings and the inclusion of individuals of all ages and employment statuses rather than just those of working age who work full-time and full-year. While both of these factors likely contribute to this result, a further explanation is that mixed race and mixed disability status households are substantially less common than mixed gender households and thus topcoding patterns are less likely to be equalized by expanding the analysis to the household. This can be observed in Table 1.3, which illustrates the average characteristics of other household members for each demographic group. From Panel A of Table 1.3, men who are topcoded live with an average of 0.60 other men and 1.38 women. Thus, the average male topcoded on his personal income leads to 1.60 total men topcoded on their household income (himself plus 0.60 other men) and 1.38 women topcoded on their household income. The income sharing within a household thus equalizes topcode rates when compared to considering all individuals separately – where a man topcoded on his own personal income would result in exactly 1 man topcoded and no women topcoded.

This is quite different than the patterns for black and disabled individuals. The average white individual who is topcoded results in 2.76 topcoded white individuals (himself or herself plus 1.76 housemates) but just 0.01 topcoded black individuals. Thus, the income sharing within households will increase, rather than decrease, the differential topcoding rates by race on household income compared to that seen for

Table 1.3: Characteristics of other household members by Gender, Race, and Disability Status

Panel 1: Composition of households by gender (1975-2008)

	All Individuals		Topcoded Individuals	
	Male	Female	Male	Female
Avg. number of additional men	1.02	1.38	0.60	1.02
Avg. number of additional women	1.49	1.01	1.38	0.50

Panel 2: Composition of households by race (1975-2008)

	All Individuals		Topcoded Individuals	
	White	Black	White	Black
Avg. number of additional white individuals	2.21	0.05	1.76	0.08
Avg. number of additional black individuals	0.01	2.68	0.01	1.75

Panel 3: Composition of households by disability status (1980-2008)

	All Individuals		Topcoded Individuals	
	Not Disabled	Disabled	Not Disabled	Disabled
Avg. number of additional non-disabled individuals	1.40	0.85	1.20	0.83
Avg. number of additional disabled individuals	0.09	0.28	0.05	0.20

Source: Author's calculations using public use March CPS data.
Note: Since household income inequality calculations include all individuals, including children, average number of additional individuals includes children as well. Panel 3 excludes children under 15, for whom disability status is not recorded.

individuals alone. Similarly for individuals with and without disabilities, the average topcoded individual without a disability results in 2.20 topcoded individuals without disabilities but just 0.05 topcoded individuals with disabilities. Thus, the income sharing within households will also increase, rather than decrease, the differential topcoding rates by disability status on household income compared to that seen for individuals alone.

1.5 Methods to correct for topcoding problems

Previous researchers have considered a range of solutions to correct for the topcoding problem. Three of these previous solutions, along with a proposed alternative solution, are discussed here. A first approach, the *Unadjusted Public Use series*, is to simply ignore topcoding issues and use the unadjusted public use CPS data as released by the Census Bureau. However, this will confuse real changes in mean income with changes in reported income due to topcoding. Additionally, as discussed previously, starting in 1995 the Census Bureau began providing a cell mean to use for all topcoded values in the public use CPS but since cell means were not provided for earlier years this major change in the reported income values among topcoded individuals results in a significant increase in measured income in 1995 and beyond (see Larrimore et al. 2008 and Burkhauser, Feng, and Jenkins, 2009). For instance, the topcode for primary earnings income rose from $99,999 to $150,000 thus reducing the share of full-time male workers who were topcoded on their own primary labor earnings from 3.93 to 1.35 percent, but the use of cell means increased the average reported primary labor earnings of those men who were still topcoded to $305,989.

A second approach, the *No Cell Mean series*, simply ignores the introduction of cell means into the public use CPS, and produces an income series where all topcoded values are assigned a value at the topcode level even after the introduction of cell means in 1995.[13] This will correct for the large artificial jump in income due to the introduction of cell means in 1995. But it does not remedy the basic problems of inconsistent topcode threshold changes over time (such as the change in primary labor earnings topcoding from $99,999 to $150,000 between 1994 and 1995) or the different topcoding rates across subgroups of the population. As a result it will still provide an inaccurate picture of income trends.

A more sophisticated approach discussed for labor earnings by Burkhauser, Butler, et al. (2004) and for income by Burkhauser, Couch, et al. (2004) is to create a

[13] A common refinement on the No Cell Mean approach is to assign topcoded individuals income that is a fixed multiple of the topcode level—e.g. 1.3 to 1.5. (See, e.g., Blau and Kahn 2000). While this comes closer to capturing levels of earnings gaps, the trends are nearly identical to those seen in the No Cell Mean series as it does not account for changes in the distribution of incomes above the topcode thresholds over time.

consistent topcode series (*Consistent Topcode series*). For each income source, this series finds the year where the topcode cuts most deeply into that source's income distribution and then chooses a topcode threshold that cuts that deeply into that source's income distribution in all other years. For example, 1.3% of individuals with wage earnings were topcoded in the public use CPS data in 2001, which is the highest percentage of individuals topcoded for wage earnings in any year. Thus, in all years an artificial topcode is imposed such that 1.3% of individuals with wage earnings are topcoded. A similar procedure was followed for all 24 sources of income to create the Consistent Topcoded Public Use series. The advantage of this approach is that all observed changes in cross-group inequality are due to real changes in the income distribution since a consistent fraction of the population is topcoded in each year. However, this consistency comes at the cost of topcoding a larger fraction of the population in all other years.

The Consistent Topcode Public Use series will consistently capture the bottom 97 percent of the earnings distribution for full-time workers and the bottom 89 percent of the household income distribution for all individuals. But, because more of the income of men, Whites, and those without work limitations is missed by consistently reducing topcodes, doing so will reduce their mean income by more than that of women, Blacks, and those with work limitations. Hence it will consistently overestimate mean income and earnings gaps between these groups. Additionally, estimates using consistent topcoding are highly dependent on the specific years of analysis chosen since the thresholds are always lowered to match the year with the most restrictive topcoding for each source. Thus, the inclusion of an additional year of data may change the inequality statistics for every year examined.

Given these limitations, this paper uses a new method for controlling for topcoding. Using the internal March CPS data files, Larrimore et al. (2008) closely followed the methodology the Census Bureau used to create cell means after 1995 and extend the series back to 1975. By using these cell means in conjunction with the public use March CPS, it is now possible to create an extended cell mean series (*Cell Mean series*) which will better match the income distributions found in the internal March CPS data for each of the demographic groups discussed in this paper.

While this approach has substantial advantages over consistent topcoding because it allows researchers to better understand changes at the top of the income distribution, it will still be impacted by changes in procedures for collecting and processing the internal CPS data. In particular, the Census Bureau performs limited censoring on internal CPS data records in addition to the more severe topcoding that occurs on the public use records. This censoring originated due to limits on the number of digits provided for responses on the survey questionnaire and early electronic data records. However, after the Census Bureau processing systems were upgraded in 1985 to allow for responses with more digits, censoring persisted out of concerns about the accuracy of these high values and the impact of such outliers on inequality statistics (See Welniak, 2003, Feng et al. 2006, and Burkhauser et al. Forthcoming for a fuller discussion of internal censoring and see Larrimore et al. 2008 for a full list of internal censoring points).

While internal censoring is a limitation of the cell mean series in measuring the "true" income trends, the censoring encountered in the internal data files used to produce the extended cell mean series is identical to that in the data files used by the Census Bureau to produce their official income statistics (Denavas-Walt, Proctor, and Smith 2009). Thus, this problem is no more severe than that which exists in these official government statistics and for researchers limited to using the public use CPS, using cell means of topcoded incomes will yield the best estimates of the US income distribution that can be made without making out of sample predictions on incomes above the internal censoring points (Burkhauser et al. Forthcoming provides one such effort to estimate censored values in the internal CPS data).

Although its less restrictive censoring and more stable censoring thresholds over time make the internal data more consistent than the public use data, there is one notable exception. Between 1992 and 1993 the Census Bureau substantially improved their data collection and processing techniques, including the implementation of computer-assisted data collection and an increase in the internal censoring thresholds (see Ryscavage 1995 and Jones and Weinberg 2000 for details on this redesign). These changes increased the ability of the Census Bureau to accurately observe incomes, particularly near the top of the distribution. While the use of internal data exacerbates the observed break in the series, changes such as computer-assisted data collection influence the entire distribution and thus may present problems for researchers using the public use files as well. Therefore, while the extended cell mean series used with the public use CPS allows for consistent trends before and after these years that closely match the internal CPS data, researchers should take caution when using any CPS based income series, including the cell means series, to compare years before 1992 to years after 1993.

1.6 Results

Impact of topcoding on mean earnings levels by demographic group. Panel A of Table 1.4 compares the trough-year ratios of mean labor earnings of men and women using four topcode correction methods to those observed in the internal March CPS data. A ratio of 1.000 indicates the correction method perfectly captures the mean earnings observed in the internal data series (Internal). The lower the ratio, the more income is missed as a result of topcoding. The four topcode correction procedures are using: the extended cell mean series together with the public use CPS data (Cell Mean), the

Table 1.4: Ratio of mean labor earnings observed in the public use CPS using alternative topcode correction methods to corresponding mean labor earnings observed using the internal CPS.

Panel 1: Mean labor earnings by gender compared to internal data

Income Year	No Cell Mean Public Use		Unadjusted Public Use		Consistent Topcode Public Use		Cell Mean Public Use	
	Female	Male	Female	Male	Female	Male	Female	Male
1975	1.000	0.986	1.000	0.986	0.998	0.950	1.000	1.000
1982	0.998	0.988	0.998	0.988	0.993	0.955	1.000	0.999
1992	0.992	0.959	0.992	0.959	0.988	0.940	1.000	1.000
1993	0.970	0.914	0.970	0.914	0.966	0.901	0.999	1.000
2004	0.974	0.934	1.002	1.003	0.966	0.907	1.002	1.003
2007	0.970	0.936	1.000	1.001	0.960	0.911	1.000	1.001

Panel 2: Mean labor earnings by race compared to internal data

Income Year	No Cell Mean Public Use		Unadjusted Public Use		Consistent Topcode Public Use		Cell Mean Public Use	
	Black	White	Black	White	Black	White	Black	White
1975	1.000	0.988	1.000	0.988	0.998	0.956	1.000	1.000
1982	0.997	0.990	0.997	0.990	0.988	0.963	1.000	0.999
1992	0.995	0.966	0.995	0.966	0.992	0.951	1.001	1.000
1993	0.962	0.927	0.962	0.927	0.959	0.916	0.996	1.000
2004	0.983	0.943	1.007	1.004	0.976	0.920	1.007	1.004
2007	0.964	0.944	1.003	1.002	0.956	0.922	1.003	1.002

Panel 3: Mean labor earnings by disability status compared to internal data[1]

Income Year	No Cell Mean Public Use		Unadjusted Public Use		Consistent Topcode Public Use		Cell Mean Public Use	
	Disabled	Not Disabled	Disabled	Not Disabled	Disabled	Not Disabled	Disabled	Not Disabled
1975								
1982								
1992	0.982	0.969	0.982	0.969	0.975	0.955	1.000	1.000
1993	0.969	0.931	0.969	0.931	0.966	0.921	0.995	0.999
2004	0.885	0.948	0.949	1.003	0.869	0.927	0.949	1.003
2007	0.919	0.948	0.949	1.001	0.908	0.929	0.949	1.001

Source: Author's calculations using internal and public use March CPS data.

[1] Disability status is not available prior to 1980 in the public use March CPS or prior to 1987 in the internal March CPS

unadjusted public use CPS data (Unadjusted Public Use), the public use CPS data without cell means (No Cell Mean), and the public use CPS data consistently topcoded (Consistent Topcode). For each series, the first column presents the ratio of mean earnings using that series to the mean earnings in the internal data for women and the second column presents the same ratio but for men.

As can be seen when looking at the data for 2007, because of Census Bureau provided cell means, the mean income of full-time, full-year male and female workers captured in the Unadjusted Public Use data since 1995 is very close to the mean values in the Internal series. So for those only interested in years since 1995, the year cell means were first provided by the Census Bureau, the Unadjusted Public Use series and the Cell Mean series both nicely capture the mean earnings of both men and women in the Internal series.

But for those also interested in years prior to 1995 the Unadjusted Public Use series is flawed because it does not provide cell means for persons above the topcoded values. Hence its mean values are smaller for both men and women. In contrast, the Cell Mean series provides yearly means very close to those from the Internal series for both men and women in all years back to 1975, coming within 0.3 percent of the internal mean values for both men and women in each of the trough years.

In contrast to these two series, both the No Cell Mean series and the Consistent Topcode series understate the mean earnings of both men and women in all years. Additionally, the amount by which earnings are understated using these series has grown over time. For example, the Consistent Topcode series understates the mean earnings seen in the internal series by 5 percent and 0.2 percent for men and women respectively in 1975. By 2007 the understatement between the Consistent Topcode and Internal series rises to 8.9 percent for male earnings and 4 percent for female earnings.

As is seen in Panels B and C of Table 1.4, similar patterns emerge for how each of the topcode corrections impact the observed mean earnings of black and white workers and of workers with and without work limitations. The group mean earnings using the Cell Mean series in all years or the Public Unadjusted series after 1995 closely match the group mean earnings in the Internal series. The group mean earnings using either the Consistent Topcode or No Cell Mean series will not only understates group mean earnings compared to the Internal series but will do so more for white that for black workers and in most years will also do so more for workers without work limitations than for workers with work limitations. However, among workers with work limitations, cell means are not consistently as accurate at capturing the internal values as they are for men and women or Blacks and Whites. This is particularly true in both 2004 and 2007 when even the Cell Mean series understates the earnings of workers with work limitations in the internal data by 5.1 percent.

Table 1.5 replicates the results presented in Table 1.4, but does so for household income of all individuals rather than labor earnings of full-time workers. When doing so, there are a few differences such as women having their mean household income understated in the No Cell Mean and Consistent Topcode series by a larger percent than their mean labor earnings were. However, while there are differences in the magnitude of how topcoding impacts income rather than earnings, the overall picture remains quite similar. Across all groups, for researchers interested in household income only after 1995, the Public Unadjusted series closely replicates the Internal mean incomes. But for researchers interested in longer term trends, prior to 1995 the Public Unadjusted series understates the mean household income in the internal data for all groups. And as was the case for labor earnings, both the No Cell Mean and Consistent Topcode series understate the mean incomes seen in the internal data for all groups in all years. The Cell Mean series, however, closely replicates the mean incomes in the internal data for all groups and thus provides a better picture of group mean incomes than that achieved using the No Cell Mean or Consistent Topcode series. Additionally, in contrast to the labor earnings series, the Cell Mean

Table 1.5: Ratio of mean household income observed in the public use CPS using alternative topcode correction methods to corresponding mean household income observed using the internal CPS.

Panel 1: Mean income by gender compared to internal data

Income	No Cell Mean Public Use		Unadjusted Public Use		Consistent Topcode Public Use		Cell Mean Public Use	
Year	Female	Male	Female	Male	Female	Male	Female	Male
1975	0.991	0.990	0.991	0.990	0.953	0.954	1.000	1.000
1982	0.993	0.993	0.993	0.993	0.958	0.960	0.999	1.000
1992	0.978	0.976	0.978	0.976	0.949	0.946	1.000	1.000
1993	0.946	0.946	0.946	0.946	0.920	0.920	0.997	1.000
2004	0.942	0.940	1.000	1.001	0.924	0.923	1.000	1.001
2007	0.939	0.941	0.999	1.003	0.924	0.926	0.999	1.003

Panel 2: Mean income by race compared to internal data

Income	No Cell Mean Public Use		Unadjusted Public Use		Consistent Topcode Public Use		Cell Mean Public Use	
Year	Black	White	Black	White	Black	White	Black	White
1975	1.000	0.989	1.000	0.989	0.988	0.949	1.000	1.000
1982	0.998	0.992	0.998	0.992	0.990	0.955	1.000	0.999
1992	0.995	0.974	0.995	0.974	0.987	0.942	1.000	1.000
1993	0.972	0.942	0.972	0.942	0.962	0.913	1.000	0.998
2004	0.977	0.938	1.003	1.004	0.971	0.917	1.003	1.004
2007	0.963	0.936	1.004	1.005	0.958	0.919	1.004	1.005

Panel 3: Mean income by disability status compared to internal data

Income	No Cell Mean Public Use		Unadjusted Public Use		Consistent Topcode Public Use		Cell Mean Public Use	
Year	Disabled	Not Disabled	Disabled	Not Disabled	Disabled	Not Disabled	Disabled	Not Disabled
1975								
1982								
1992	0.996	0.977	0.996	0.977	0.956	0.946	1.001	1.000
1993	0.991	0.948	0.991	0.948	0.958	0.920	1.002	1.000
2004	0.955	0.941	0.994	1.000	0.939	0.923	0.994	1.000
2007	0.956	0.940	0.996	1.001	0.948	0.926	0.996	1.001

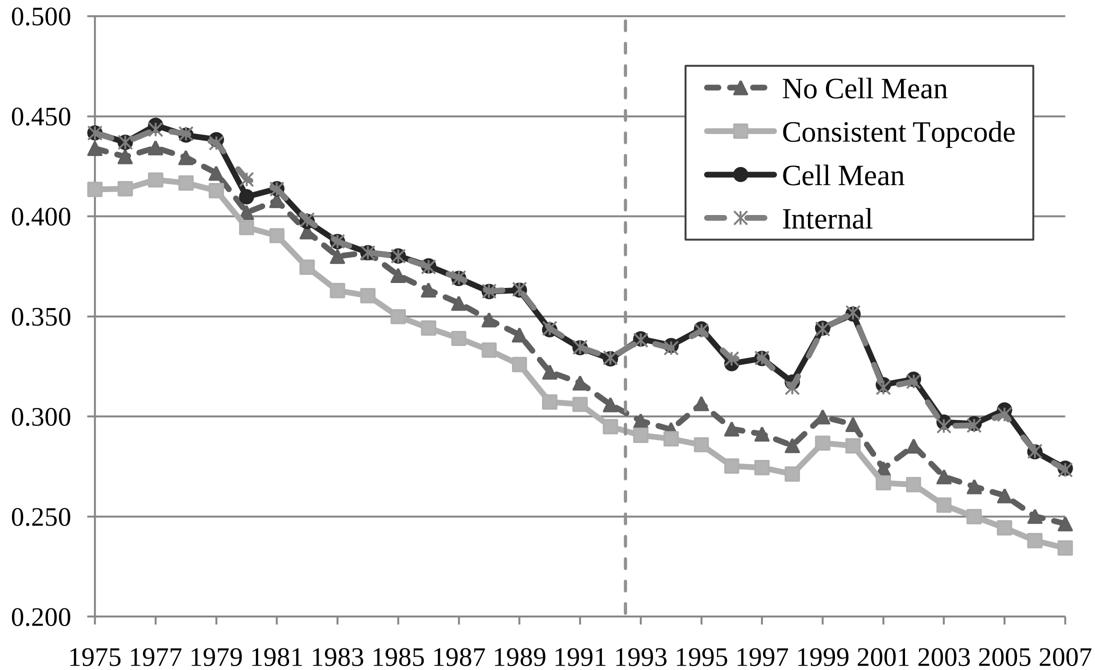
Source: Author's calculations using internal and public use March CPS data.
[1] Disability status is not available prior to 1980 in the public use March CPS or prior to 1987 in the internal March CPS. Comparisons of household income for people with and without a disability excludes individuals age 15 and under, for whom disability status is not captured in the March CPS

Series for household income closely replicates the mean values in the internal data for workers with disabilities in all years of the series including both 2004 and 2007 when it is within 0.6 percent of the internal values.

Impact of topcoding on mean earnings gaps. Having observed that mean incomes and earnings of all groups are influenced by topcoding, how do these differences in mean income and earnings levels influence mean income and earnings gaps? In addressing these questions, this paper focuses on differences in the No Cell Mean, Consistent Topcode, Cell Mean and Internal series. The Unadjusted Public Use series is excluded from further discussions since it is identical to the No Cell Mean series before 1995 and after 1995 it is nearly identical to the Cell Means series. Since the Unadjusted Public Use series is a combination of these two series, it cannot provide additional information about trends in the earnings gaps, and has a clear artificial jump in 1995 that makes it inferior to either of its component series individually.

Because the No Cell Mean and Consistent Topcode series consistently understate the labor earnings of both men and women prior to the introduction of cell means, the male-female mean earnings gap could in principal be greater or less than that in the Cell Mean and Internal series. But as shown in Tables 1.1 and 1.4, men are more likely than women to have labor earnings topcoded and their amount of suppressed earnings from topcoding is greater. Therefore, the relative earnings of women will appear to be higher and the earnings gap lower in the No Cell Mean and Consistent Topcode series than in the Cell Mean and Internal series.

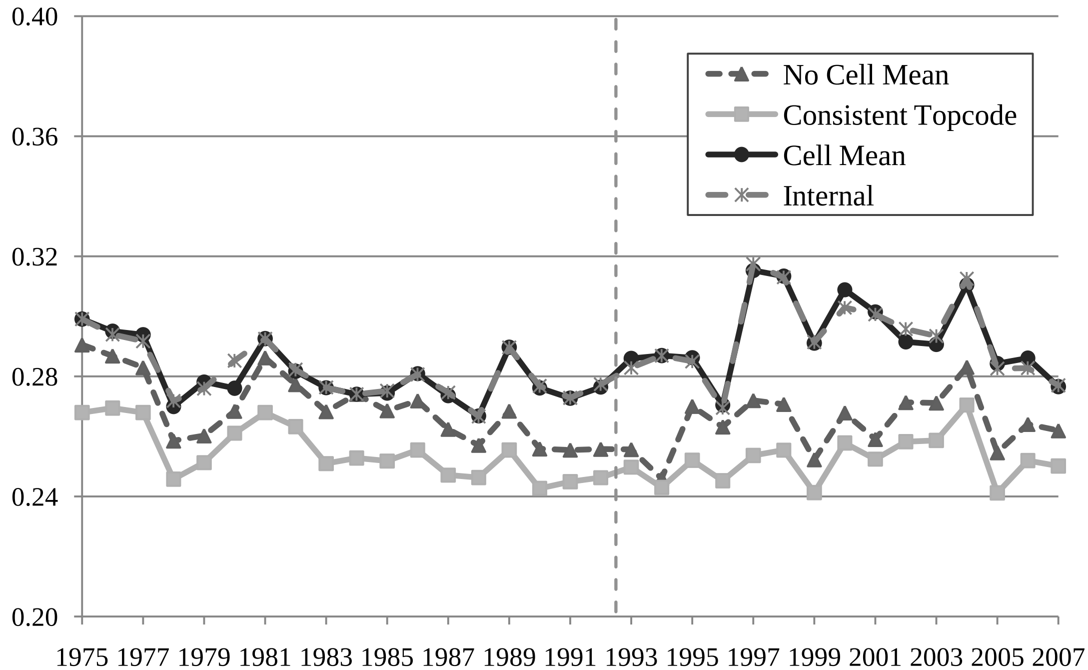
This can be seen in Figure 1.1 which compares the gender mean earnings gap for full-time workers of working age in each of the four series. In all years, the female-male earnings gap is smaller using the No Cell Mean series than the Internal series. This difference is relatively small in the first year of the sample, but has grown over time. In 1975 it was under 1 percentage point—females' earned 43.4 percent less than male workers' using the No Cell Mean Public Use Series and 44.2 percent using the



Source: Author's calculations using internal and public use March CPS data.
Figure 1.1: Mean labor earnings gap between men and women using alternative topcode correction methods

Internal series—but the difference between these series grew to over 2 percentage points by 1989 and was 2.7 percent in 2007.

This understatement is even greater when Consistent Topcoding is used, since it further suppresses values at the top of the earnings distribution and captures even more male earnings relative to female earnings. Using Consistent Topcoding understates the gap between female and male mean earnings by 2.8 percentage points in 1975 and this understatement rises to 3.9 percentage points by 2007. The growth in the understatement of the gender earnings gap is important for evaluating the trends in the relative earnings of women. Since the extent to which both the Consistent Topcode series and the No Cell Mean series understate the gender earnings gap has grown over time, using either of these series will overstate the relative improvement in female earnings in addition to overstating their relative earnings in each year.

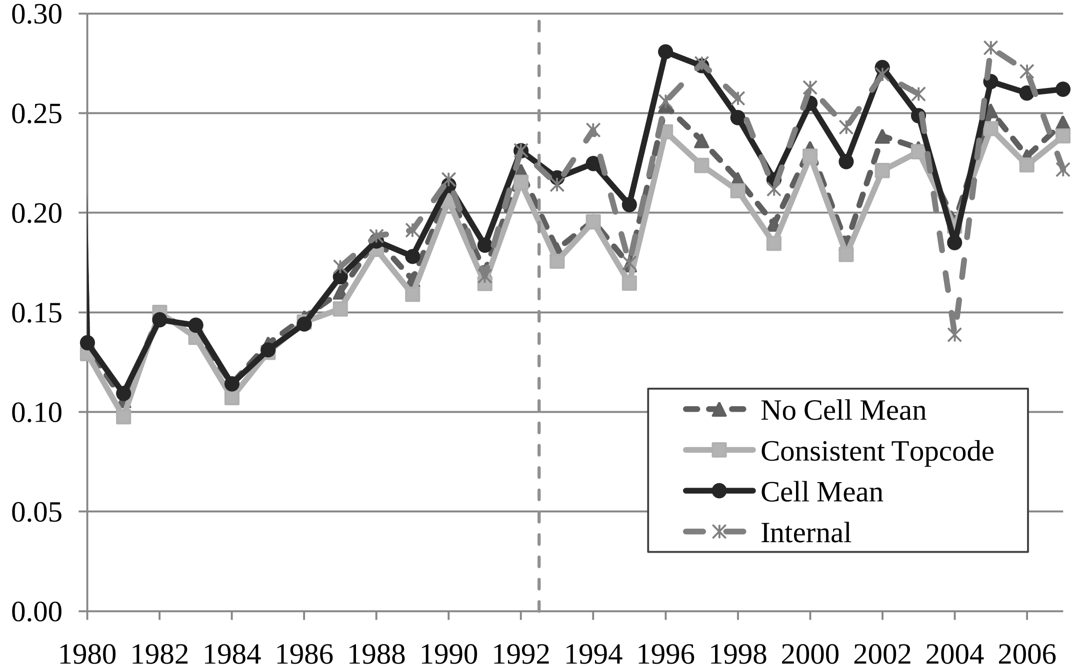


Source: Author's calculations using internal and public use March CPS data.
Figure 1.2: Mean labor earnings gap between blacks and whites using alternative topcode correction methods

In contrast to the Consistent Topcode and No Cell Mean series, the Cell Mean series nicely approximates the gender earnings gap found using the Internal CPS data in all years. Thus, unlike that seen for the Consistent Topcode and No Cell Mean series, it also closely approximates the trend in the female relative earnings since 1975. Therefore, for researchers limited to the Public Use data, the Cell Mean series is better able to replicate internal results from the Internal data than previous topcode correction techniques.

Figure 1.2 provides a similar analysis of the Black-White earnings gap under each of the four methods of controlling for topcoding.[14] As was seen for the female-male earnings gap, using the No Cell Mean series understates the racial earnings gap

[14] Since earnings and income gaps are only capable of comparing two groups, Hispanics and individuals of races other than white or black are excluded in this analysis. These individuals are included, however, in the analysis of within-group and between-group inequality later in the paper.



Source: Author's calculations using internal and public use March CPS data.
Note: Disability status is not available prior to 1980 in the public use March CPS or prior to 1987 in the internal March CPS

Figure 1.3: Mean labor earnings gap between people with and without a disability using alternative topcode correction methods

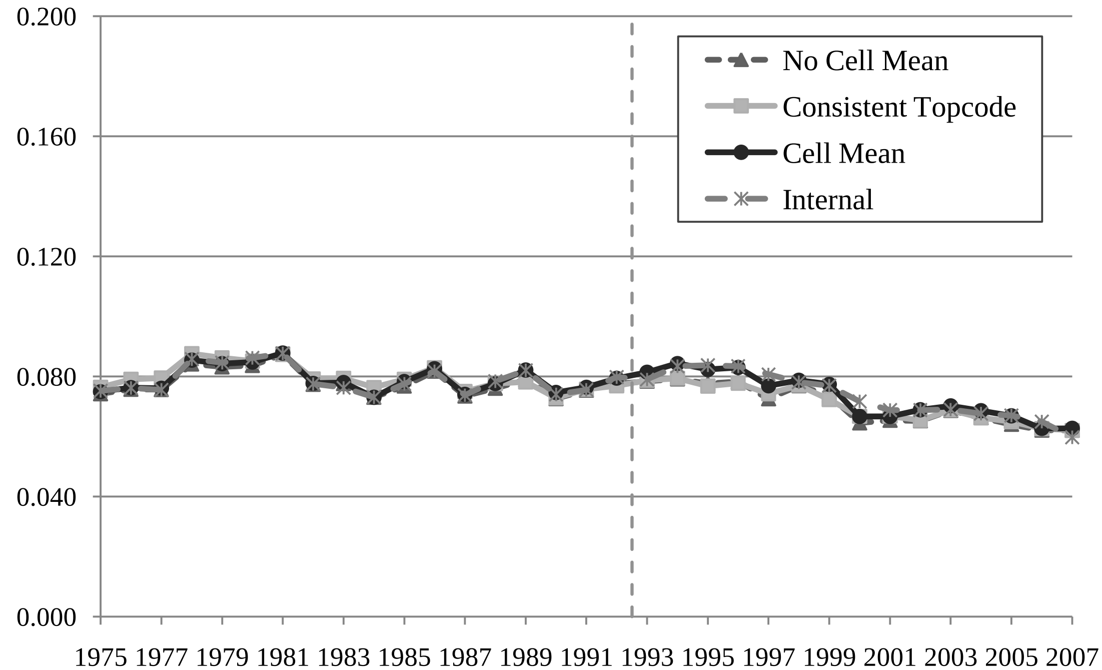
and the extent of this understatement grew over time from 0.8 percentage points in 1975 to 2.7 percentage points in 2004 before falling back to 1.5 percentage points in 2007. Similarly, the Consistent Topcode series understates the black-white earnings gap by even more, as white workers are more likely to be near the top of the earnings distribution and thus have additional earnings suppressed by using this topcode correction method. Once again, however, the Cell Means Public Use series closely matches the earnings gaps from the Internal CPS data and is the best available method of replicating the earnings gap seen in the Internal series.

For workers with and without disabilities (Figure 1.3), the earnings gap among full-time, full-year workers is smaller than that seen across races or genders. This may be expected as workers with relatively mild disabilities may be more likely to work full-time so the selection effect reduces the earnings gap. Additionally, the small sample of full-time workers with disabilities and the selection effects into this group result in some ambiguity in the magnitude and direction of topcoding's effect on the disability earnings gap. In general both the Consistent Topcode and No Cell Mean series overstate the earnings gaps seen in the internal data, although this is less consistent over time than was the case for the gender and racial earnings gaps. Once again, however, the Cell Mean series is superior in its ability to replicates the internal results when compared to these previous approaches used with the public use data.

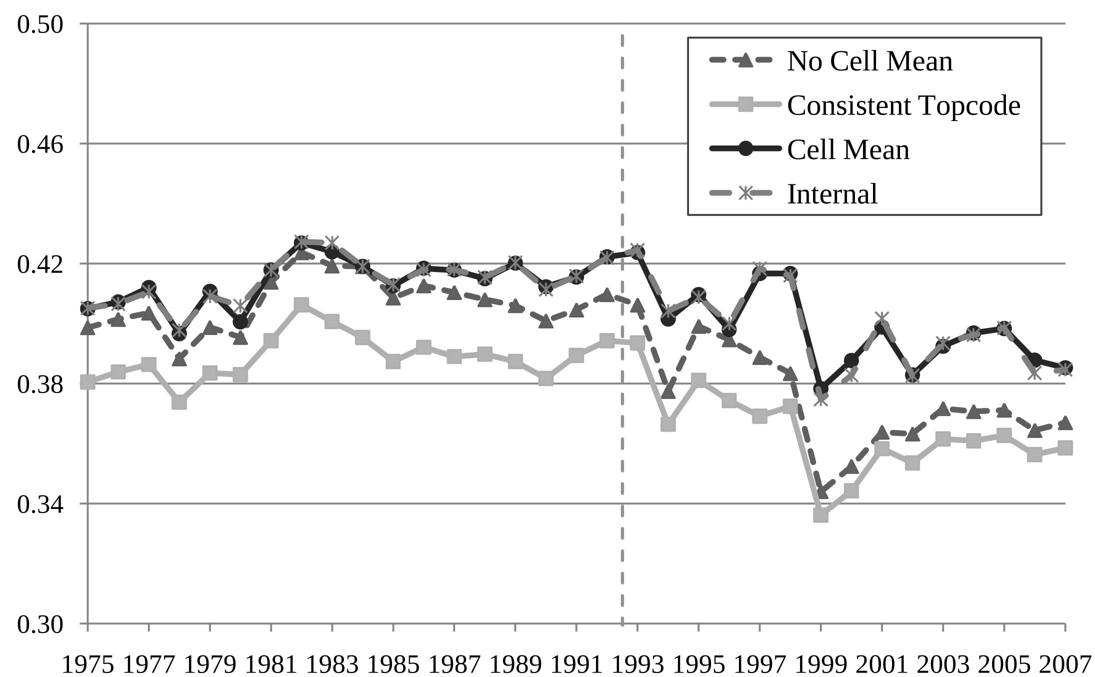
Impact of topcoding on mean household income gaps. It is apparent above that the levels and trends in mean earnings gaps are impacted by the choice of topcode correction method. This effect is also apparent in the gaps in household income, which includes both non-labor income and the sharing of resources between household members. This can be seen in Figures 1.4, 1.5 and 1.6, which illustrate the mean household income gaps by race, gender, and disability status respectively.

A comparison of the gender income gap to the gender earnings gaps found previously reveals that focusing on household income reduces the size of the gap. For example, in each of the four series examined, the gender income gap of approximately 6 percent in 2007 is well smaller than the 23 to 28 percent gender earnings gaps seen that year. Additionally, the impact of topcoding on the gender income gap is much smaller than that for earnings gaps. The difference between the gender income gaps using each of the four topcode correction methods is never more than 0.7 percent and is less than 0.2 percent in many years, which is well below the differentials of up to 6 percent seen for labor earnings gaps.

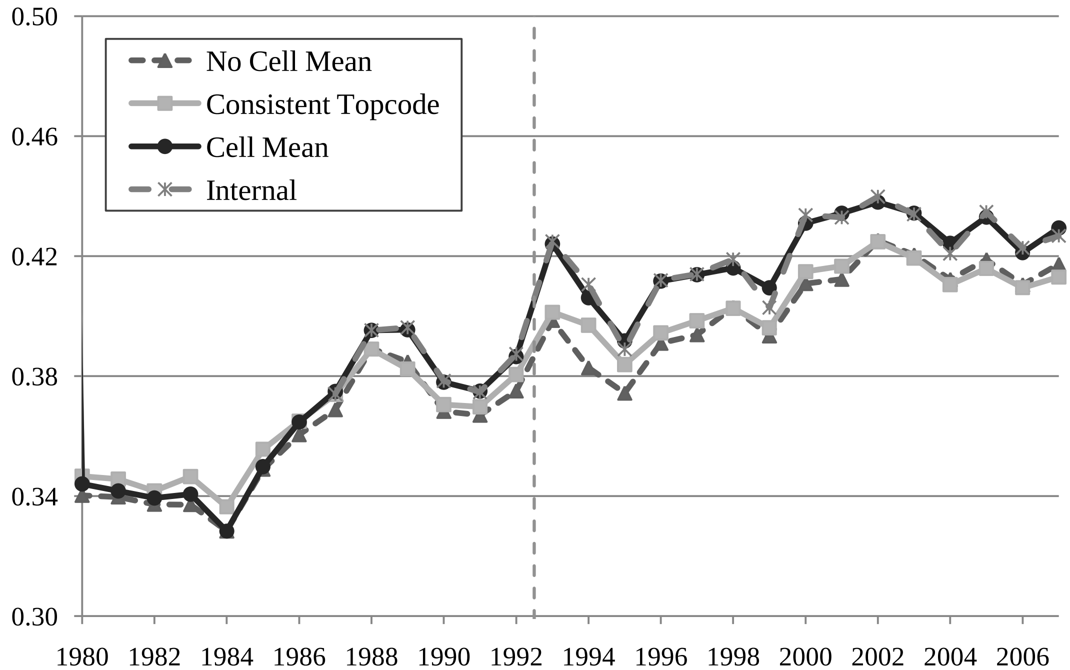
While the gender income gap is substantially smaller than the gender earnings gap, the same is not true for the income and earnings gaps by race and disability



Source: Author’s calculations using internal and public use March CPS data.
Figure 1.4: Mean household income gap between men and women using alternative topcode correction methods



Source: Author’s calculations using internal and public use March CPS data.
Figure 1.5: Mean household income gap between Blacks and Whites using alternative topcode correction methods



Source: Author's calculations using internal and public use March CPS data.
Note: Disability status is not available prior to 1980 in the public use March CPS or prior to 1987 in the internal March CPS

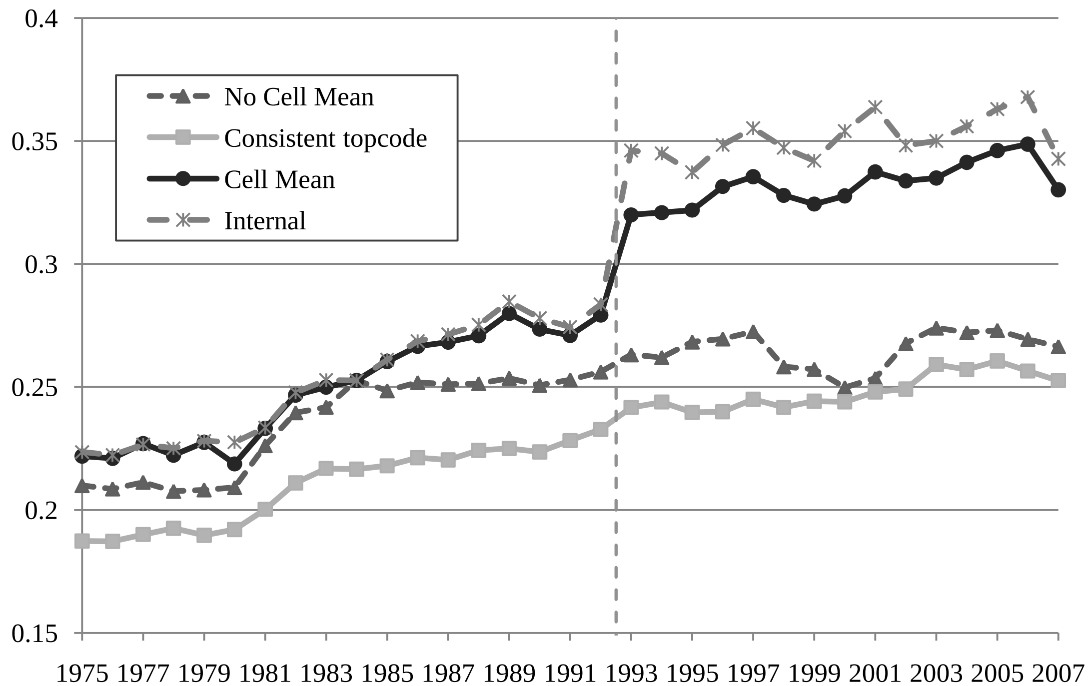
Figure 1.6: Mean household income gap between people with and without a disability using alternative topcode correction methods

status. Instead, a comparison of the racial income gap in Figure 1.5 with the racial earnings gap in Figure 1.2 reveals that the income gap between Blacks and Whites is larger than the earnings gap between full-time black and white workers. Additionally, as was seen for the racial earnings gaps, using the No Cell Mean or Consistent Topcode series understates the racial income gap seen in the Internal series while the Cell Mean series closely approximates these internal results. Similarly, a comparison of the disability income gap in Figure 1.6 with the disability earnings gap in Figure 1.3 reveals that the income gap between individuals with and without a disability is larger than the earnings gap between full-time workers with and without disabilities. And comparing the income gaps across the 4 topcode series, the income gap between disabled and non-disabled workers is understated in the Consistent Topcode and No

Cell Mean series, mirroring the results found for each of the other income and earnings gaps calculated.

The same explanations discussed previously for how relative topcoding rates differ for labor earnings and household income may apply here as well when considering the different earnings and income gaps across the three sets of demographic groups. The relatively high prevalence of mixed gender households, as previously illustrated in Table 1.3, reduces both the gender income gap and the impact of topcoding on this gap. Since mixed race and mixed disability status households are less common, they should have less of an impact on the race or disability status income gaps and may even contribute to the widening of these income gaps compared to the corresponding earnings gaps. Additionally, the larger household income gap by disability status than that seen for labor earnings may be partially attributable to the fact that the income gap incorporates individuals including those who do not work – including more severely disabled individuals – whereas the earnings gap is only calculated for individuals working full-time. As a result, any selection effects into the labor market by severity of disabilities and earnings potential will reduce the disability earnings gap but not the disability income gap. The broadened age range, in addition to the selection effects into the labor market, may also contribute to this result in a similar fashion. Finally, the other differences between household income and labor earnings – including different rates of receipt of non-labor earnings by group, the prevalence of marriage and large or small households within each group, and differences in the correlation of earnings within a household by group – could also lead to the observed differences between the household income gaps and those seen for labor earnings.

Decomposing population-wide household income inequality by demographic group. Given that the compression of the income distribution from topcoding



Source: Author's calculations using internal and public use March CPS data.
Figure 1.7: Theil household income inequality index using alternative topcode correction methods

influenced cross-group inequality as measured by mean income gaps, it should not be surprising that it also influences population-wide income inequality statistics. This is apparent in Figure 1.7, which displays population-wide household income inequality calculated using the Theil inequality index for each topcode correction method (See Appendix Figures 1.1 and 1.2 for results using the relatively bottom-sensitive GE(0) index and the relatively top-sensitive GE(2) index).[15] When using the No Cell Mean series or the Consistent Topcode series, the observed level of inequality well understates the inequality observed in the Internal series using the Theil index. Additionally, the No Cell Mean series misstates trends in inequality from changes in the topcode threshold. Most recently, this is evident between 2001 and 2002 when

[15] The Generalized Entropy family of inequality indices cannot be computed using zero or negative incomes. Therefore, as is common for discussions of these inequality measures, when computing the Theil and other Generalized Entropy indices these incomes are bottomcoded at $1 (see Burkhauser et al. Forthcoming and Jenkins et al. Forthcoming for a further discussion of this approach).

inequality in the No Cell Mean series increased as the topcode threshold increased but declined slightly in the more accurate Internal and Cell Mean series. Similar effects occurred with previous changes to the topcode threshold as well.

In contrast, the Cell Mean series which accurately captures the level, if not the dispersion, of top incomes more closely matches both the level and trends of income inequality in the Internal Series. However, while this dispersion of top incomes generally did not impact mean income gaps, which were virtually identical using the Internal data and the Cell Mean series, the lack of dispersion in the Cell Mean series does have a noticeable effect on the Theil index. In order to remove the remaining difference between the Cell Mean and Internal data, it is necessary to also include information about the dispersion of top incomes in the topcode correction mechanism (see Burkhauser, Feng, and Larrimore, Forthcoming, for information on the variance of top incomes and one approach for incorporating this data into the topcode correction).[16] Nevertheless, with the exception of the break between 1992 and 1993 when the Internal data shows a larger jump in inequality than the Cell Mean series, the trend in inequality between the Cell Mean and Internal series are quite similar. Additionally, while not a perfect fit, the level of population-wide inequality using the Cell Mean series is substantially closer to that in the Internal data than in either the No Cell Mean or Consistent Topcode series.

A valuable property of the Theil index, and other Generalized Entropy indices, is that unlike the Gini coefficient it is additively decomposable by subgroups. Thus, it can be observed whether the higher income inequality in the Internal and Cell Mean series comes from within or between the demographic groups described above, as well

[16] As can be seen in Appendix Figures 1.1 and 1.2, for researchers particularly interested in top-sensitive inequality indices, the gap between the level of inequality in the Cell Mean series and the Internal series increases, while for researchers interested in bottom-sensitive inequality indices the difference in levels declines.

as which subgroups are most important for understanding the higher inequality in these series. Unlike mean income gaps which only permit a comparison between two groups, the Theil decomposition allows for multiple groups to be considered at once. While this is not necessary for naturally binary groups such as gender and disability status, it allows more delineation across races than was possible in the black-white income gap. Thus, when decomposing the household income inequality by race, four racial groups will be considered: white, black, Hispanic, and all other races.

The three panels of Table 1.6 illustrate the results of decomposing the Theil household income inequality index for people of all ages by gender, race, and disability status in each of the trough years since 1975 (inequality decompositions for the alternative relatively bottom-sensitive and relatively top-sensitive inequality measures, GE(0) and GE(2), are available in Appendix Tables 1.4 and 1.5).[17] When decomposing the population-wide inequality by any of these demographic characteristics, it is apparent that the vast majority of household income inequality exists within these demographic groups rather than between groups. In each trough year examined, less than 1 percent of household income inequality using each topcode series is between genders (for example, 0.001 of the total within and between inequality of 0.336 in 2004 using internal data), less than 5 percent is between disability statuses (for example, 0.010 out of the total within and between inequality of 0.346 in 2004 using the internal data), and less than 8 percent is between races. In the sections above it was illustrated that correcting for topcoding using Cell Means or Internal series rather than using the No Cell Mean or Consistent Topcode series finds greater income gaps between each of these demographic groups. While this is reflected in the Theil decomposition through a higher between-group inequality

[17] As was the case above for household income gaps, individuals under age 15 are excluded when considering inequality within and between groups of individuals based on disability status since disability status is only asked of people over age 15.

Table 1.6: Decomposition of Theil household income inequality into within-group and between-group inequality using alternative topcode correction methods

Panel A: Decomposition of Theil household income inequality by gender

	No Cell Mean				Consistent Topcode				Cell Mean				Internal			
Income Year	Within Female	Within Male	Total Within	Total Between	Within Female	Within Male	Total Within	Total Between	Within Female	Within Male	Total Within	Total Between	Within Female	Within Male	Total Within	Total Between
1975	0.217	0.201	0.209	0.001	0.195	0.201	0.209	0.001	0.229	0.214	0.221	0.001	0.230	0.215	0.223	0.001
1982	0.247	0.231	0.239	0.001	0.218	0.231	0.239	0.001	0.254	0.238	0.246	0.001	0.255	0.239	0.247	0.001
1992	0.266	0.245	0.255	0.001	0.243	0.245	0.255	0.001	0.288	0.269	0.278	0.001	0.292	0.274	0.283	0.001
1993	0.274	0.251	0.262	0.001	0.253	0.251	0.262	0.001	0.331	0.308	0.319	0.001	0.358	0.333	0.345	0.001
2004	0.283	0.264	0.273	0.001	0.268	0.264	0.273	0.001	0.344	0.325	0.334	0.001	0.365	0.346	0.355	0.001
2007	0.276	0.256	0.266	0.001	0.263	0.256	0.266	0.001	0.341	0.319	0.330	0.001	0.355	0.329	0.342	0.000

Panel B: Decomposition of Theil household income inequality by race

	No Cell Mean						Consistent Topcode					
Income Year	Within Black	Within White	Within Hispanic	Within Other	Total Within	Total Between	Within Black	Within White	Within Hispanic	Within Other	Total Within	Total Between
1975	0.249	0.191	0.224	0.222	0.197	0.013	0.253	0.167	0.213	0.201	0.175	0.012
1982	0.292	0.215	0.269	0.243	0.224	0.016	0.284	0.185	0.254	0.222	0.197	0.014
1992	0.343	0.223	0.292	0.261	0.238	0.018	0.335	0.198	0.278	0.242	0.216	0.017
1993	0.352	0.227	0.293	0.270	0.243	0.020	0.342	0.205	0.279	0.253	0.223	0.018
2004	0.324	0.240	0.289	0.284	0.255	0.019	0.317	0.224	0.282	0.269	0.241	0.018
2007	0.320	0.232	0.279	0.265	0.247	0.020	0.313	0.217	0.270	0.251	0.234	0.019

	Cell Mean						Internal					
Income Year	Within Black	Within White	Within Hispanic	Within Other	Total Within	Total Between	Within Black	Within White	Within Hispanic	Within Other	Total Within	Total Between
1975	0.249	0.204	0.231	0.229	0.208	0.014	0.249	0.205	0.231	0.225	0.210	0.014
1982	0.295	0.222	0.275	0.251	0.231	0.016	0.295	0.223	0.276	0.252	0.232	0.016
1992	0.350	0.247	0.304	0.285	0.260	0.019	0.351	0.252	0.304	0.294	0.264	0.019
1993	0.394	0.285	0.333	0.330	0.298	0.022	0.404	0.313	0.343	0.371	0.324	0.022
2004	0.364	0.303	0.333	0.344	0.313	0.022	0.366	0.328	0.353	0.349	0.335	0.021
2007	0.381	0.296	0.319	0.325	0.308	0.022	0.387	0.311	0.325	0.343	0.321	0.022

Table 1.6 (continued)

Panel C: Decomposition of Theil household income inequality by disability status[1]

	No Cell Mean				Consistent Topcode				Cell Mean				Internal			
Income Year	Within Disabled	Within Not Disabled	Total Within	Total Between	Within Disabled	Within Not Disabled	Total Within	Total Between	Within Disabled	Within Not Disabled	Total Within	Total Between	Within Disabled	Within Not Disabled	Total Within	Total Between
1982	0.286	0.221	0.225	0.006	0.247	0.192	0.196	0.006	0.291	0.228	0.232	0.006				
1992	0.307	0.234	0.239	0.008	0.269	0.211	0.214	0.008	0.315	0.257	0.261	0.008	0.313	0.262	0.265	0.008
1993	0.297	0.239	0.243	0.010	0.264	0.217	0.220	0.010	0.312	0.295	0.296	0.011	0.312	0.318	0.318	0.011
2004	0.303	0.253	0.256	0.010	0.287	0.239	0.242	0.010	0.343	0.312	0.314	0.011	0.380	0.334	0.336	0.010
2007	0.306	0.246	0.249	0.009	0.297	0.232	0.236	0.009	0.353	0.306	0.309	0.010	0.370	0.319	0.322	0.010

Source: Author's calculations using Internal and Public Use March CPS data.

[1]Disability status is not available prior to 1980. Comparisons of household income for people with and without a disability excludes individuals age 15 and under, for whom disability status is not captured in the March CPS

component, given the small impacts that the inequality between these groups have on total inequality, the actual increase in between-group inequality from improving the topcode correction method is quite small. When decomposing by race, where the between-group inequality is most important to household income inequality, in 2004 using the Internal series rather than the No Cell Mean series increases the observed between-group component of the Theil index by 0.002. Yet, this is just a small element of the 0.082 increase in the observed population-level Theil index in 2004 when the Internal series is used rather than the No Cell Mean series. Using the Internal series has an even smaller effect on the between-group component of inequality when decomposing by gender or disability status – in 2004 less than 0.0001 of the 0.082 increase in the population's Theil index switching from the No Cell Mean series to the Internal series came from the between-gender component of inequality.

Given that much of the household income inequality comes from within groups, it is valuable to also observe which groups have the highest levels of within-group income inequality. Perhaps surprisingly, in almost all years the household income inequality is greater among women than it is among men; among Blacks, Hispanics, and other minority races that it is among Whites; and among individuals with a disability than it is among individuals without a disability. However, the magnitude of these differences in within-group inequality varies based on the topcode correction technique used.

Since the Consistent Topcode and No Cell Mean series suppress incomes for individuals at the top of the distribution in all groups, using either of these series reduces the within-group inequality in each of the demographic groups. However, the amount by which it reduces within-group inequality varies by group. For each gender, the extent to which within-group income inequality is suppressed by the Consistent Topcode or No Cell Mean series is similar in any given year. For example, in 2004, which is the last trough year in the data, using the No Cell Mean Series rather than the Cell Mean series reduces within-group Theil index for male inequality by 0.061 (0.264 versus 0.325) and reduces within-group female inequality by the same amount (0.283 versus 0.344). The same is true when comparing the No Cell Mean series to the Internal series, as the within-group Theil index for both male and female inequality is 0.082 lower in the No Cell Mean series than in the Internal one. Thus, while censoring impacts the levels and trends in within-group inequality for both men and women it does not alter the relative within-group inequality for the genders.

This relationship is not true for comparisons of inequality within different racial groups or within groups of different disability statuses. Household income inequality is always the highest among black individuals. But since less Blacks are subject to topcoding than Whites, using the No Cell Mean or Consistent Topcode series understates the inequality among Blacks by less than it does Whites. Thus, when using the more accurate Cell Mean or Internal series the difference in within-group inequality for black and white individuals declines. In 2004 using the Consistent Topcode series finds that the Theil index among Blacks is 0.092 higher than among Whites and using the No Cell Mean series it is 0.084 higher. But when using the Cell Mean series this difference in Theil indices shrinks to 0.061 and when using the internal series it shrinks even further to 0.038. A similar relationship exists when comparing the within-group inequality of other races as well. While inequality among Hispanics is always greater than inequality among Whites, the magnitude of this difference shrinks once correcting for topcoding using the Cell Mean Series rather than the No Cell Mean or Consistent Topcode series.

As was the case for racial groups, the difference between within-group inequality for individuals with disabilities and that for individuals without disabilities declines somewhat from better capturing high incomes. Household income inequality is always the highest among individuals with disabilities. But since less of these individuals are subject to topcoding than individuals without disabilities, using the No Cell Mean or Consistent Topcode series understates the inequality among individuals with disabilities less than it does those without. Thus, when using the more accurate Cell Mean series or Internal series the difference in within-group inequality for individuals with and without disabilities declines.

1.7 Conclusions

Topcoding is a well-documented problem for the Current Population Survey, but until recently, the only available strategy for mitigating the problem has been to place further restrictions on the data, either by using consistent topcoding or by discarding the cell means provided by the Census Bureau from 1995 onward. Each of these previous topcode correction methods relied on gaining consistency by reducing the usable information about top incomes.

Through access to internal March CPS data, the constraints of topcoding have largely been lifted which allows for a more accurate view of inequality in the United States for the complete income distribution. Unlike previous topcode correction methods, the cell mean series created based off of this internal data, and made available to the public in Larrimore et al. (2008), gains consistency by increasing the usable information about top incomes. This paper demonstrated that within demographic group, between demographic groups, and population-wide inequality statistics are all impacted by the topcode correction method. In each case the results using cell means more closely mirror results based on the more complete internal March CPS data.

When using the internal March CPS data or public use data with cell means, earnings and income inequality within-group and between-group are higher than that seen using previously available topcode correction methods. Measuring between-group earnings inequality first using mean earnings gaps, the mean earnings gap between men and women, Blacks and Whites, and people with and without disabilities are higher than that seen using the public use data without cell means. The same is true for household income gaps across race and disability status. For gender, however, there is only a small household income gap across genders and similarly only a small effect from topcoding.

In addition to understating between-group income inequality, previous topcode corrections have also led to an understatement of population-wide household income inequality. Thus, once correcting for topcoding using internal data or cell means, it is apparent that the level of inequality for the population is higher than that observed in the public use data and has been higher for the past 40 years. While the greater between-group income inequality slightly contributed to this finding, the primary reason for the higher population-wide inequality using cell means or the internal data comes through a rise in within-group inequality. The vast majority of population-wide household income inequality is within, rather than between, gender, racial, and disability status demographic groups. Once using cell means to correct for topcoding, the within-group inequality increases substantially which results in the greater population-wide income inequality.

Finally, topcoding also impacts which demographic groups have higher within-group inequality. In particular, while Blacks and Hispanics have greater within-group inequality than Whites in the no cell mean series the magnitude by which inequality is greater among Blacks than Whites declines when using cell means and, in recent years, inequality is greater among Whites than Hispanics when using cell means. This observation that topcoding impacts each demographic group differently emphasizes the need to carefully consider the treatment of top incomes when calculating inequality statistics. While one may expect that previous topcode correction techniques reduce inequality and impact its trends as topcoding thresholds vary, the intricacies of who is topcoded in any given year makes the choice of topcode correction at least as important for such comparisons of inequality within and between demographic groups as well.

APPENDIX

Appendix Table 1.1: Number of individuals by demographic group in the March Current Population Survey

Panel A: All Individuals						
Income Year	Male	Female	White	Black	Not Disabled	Disabled
1975	64,149	68,989	105,203	13,725		
1976	76,166	81,484	124,487	15,540		
1977	73,815	79,192	120,199	15,098		
1978	73,110	78,874	118,474	15,080		
1979	86,156	92,186	139,259	16,973		
1980	86,083	92,313	138,606	17,033	122,690	13,125
1981	77,169	82,849	123,485	15,592	109,755	12,417
1982	77,093	82,797	123,288	15,634	110,706	11,811
1983	76,145	82,307	121,946	15,312	109,221	12,238
1984	76,108	82,642	121,560	15,789	109,621	12,479
1985	74,514	80,370	118,848	15,306	107,778	11,744
1986	73,457	79,517	116,817	15,538	106,825	11,261
1987	73,809	79,558	116,961	15,151	107,573	10,849
1988	68,608	73,700	109,838	13,796	99,835	10,142
1989	74,838	80,678	116,121	15,301	108,647	10,992
1990	75,008	80,811	114,982	14,997	108,498	10,995
1991	73,951	79,519	112,326	15,109	107,087	10,895
1992	73,555	79,499	111,673	15,000	106,009	11,295
1993	71,253	77,501	108,251	14,708	101,383	12,150
1994	70,655	76,816	105,919	14,526	100,934	12,021
1995	61,569	67,240	91,264	12,702	88,050	10,504
1996	62,533	67,586	91,522	12,860	89,536	10,534
1997	62,733	67,353	90,997	12,524	90,192	9,821
1998	63,128	67,726	90,954	12,711	91,030	9,707
1999	63,989	68,207	89,976	12,735	92,254	9,930
2000	61,874	65,506	85,791	12,318	89,288	9,553
2001	103,935	110,588	147,439	24,676	145,398	14,271
2002	103,967	109,817	143,403	24,707	145,512	14,027
2003	101,920	108,544	139,809	23,934	143,486	14,294
2004	100,805	107,120	137,323	23,429	141,743	14,222
2005	99,826	106,018	133,199	23,353	140,838	14,466
2006	99,289	104,884	131,377	23,617	140,875	13,410
2007	98,655	105,117	129,957	23,896	141,464	13,375

Appendix Table 1.1 (Continued)

Panel B: Working age, full time workers						
Income Year	Male	Female	White	Black	Not Disabled	Disabled
1975	21,235	9,876	25,993	2,471		
1976	25,614	11,860	31,247	2,933		
1977	25,483	12,331	31,330	3,002		
1978	25,954	13,072	32,132	3,062		
1979	30,906	15,954	38,634	3,543		
1980	30,276	16,279	38,119	3,499	45,606	949
1981	26,893	14,880	33,879	3,226	40,903	870
1982	25,735	15,166	33,332	3,099	40,111	790
1983	26,210	15,809	34,017	3,289	41,164	855
1984	27,535	16,679	35,597	3,553	43,328	886
1985	27,289	16,699	35,318	3,690	43,093	895
1986	27,148	17,056	35,351	3,723	43,331	873
1987	27,648	17,776	36,278	3,706	44,671	753
1988	26,058	16,966	34,703	3,456	42,312	712
1989	28,491	18,418	36,659	3,896	46,092	817
1990	27,826	18,491	35,886	3,713	45,530	787
1991	26,757	18,645	35,086	3,685	44,606	796
1992	26,510	18,790	34,990	3,638	44,513	787
1993	26,012	18,196	34,002	3,593	43,437	771
1994	26,480	18,399	34,149	3,732	44,113	766
1995	23,367	16,390	29,843	3,333	39,051	706
1996	23,859	16,686	30,085	3,410	39,855	690
1997	24,323	17,119	30,391	3,509	40,813	629
1998	25,103	17,567	31,049	3,639	42,040	630
1999	25,539	18,176	31,027	3,852	43,019	696
2000	24,855	17,854	29,907	3,846	42,078	631
2001	40,109	29,212	49,174	7,332	68,355	966
2002	39,475	28,636	47,631	7,097	67,252	859
2003	38,602	28,148	46,237	6,834	65,911	839
2004	38,412	27,941	45,502	6,746	65,525	828
2005	38,577	28,000	44,843	6,744	65,736	841
2006	38,714	28,207	44,527	7,127	66,204	717
2007	38,268	28,651	44,328	7,224	66,227	692

Source: Author's calculations using public use March CPS data.
Note: Disability status is not available prior to 1980 in the public use March CPS or prior to 1987 in the internal March CPS

Appendix Table 1.2: Public use CPS censoring points for each income source in dollars (1975-1986)

	Wages (I51A)	Self Employment (I51B)	Farm (I51C)	Social Security (I52A)	Supplemental Security (I52B)	Public Assistance (I53A)	Interest (I53B)	Dividends Rentals (I53C)	Veterans/ Workers Comp (I53D)	Retirement (I53E)	Other (I53F)
1975-1980	50,000	50,000	50,000	9,999	5,999	19,999	50,000	50,000	29,999	50,000	50,000
1981-1983	75,000	75,000	75,000	19,999	5,999	19,999	75,000	75,000	29,999	75,000	75,000
1984-1986	99,999	99,999	99,999	19,999	9,999	19,999	99,999	99,999	29,999	99,999	99,999

Source: Current Population Survey Annual Demographic File Technical Documentation

Appendix Table 1.3: Public Use CPS Censoring Points for each Income Source in Dollars (1987–2007)

Year	Primary Earnings (ERN_VAL)	Wages (WS_VAL)	Self Employment (SE_VAL)	Farm (FRM_VAL)	Social Security (SS_VAL)	Supplemental Security (SSI_VAL)	Public Assistance (PAW_VAL)	Interest (INT_VAL)
1987-1992	99,999	99,999	99,999	99,999	29,999	9,999	19,999	99,999
1993-1994	99,999	99,999	99,999	99,999	29,999	9,999	19,999	99,999
1995-1997	150,000	25,000	40,000	25,000	49,999	25,000	24,999	99,999
1998-2001	150,000	25,000	40,000	25,000	49,999	25,000	24,999	35,000
2002-2007	200,000	35,000	50,000	25,000	49,999	25,000	24,999	25,000

Year	Dividends (DIV_VAL)	Rental (RNT_VAL)	Alimony (ALM_VAL)	Child Support (CSP_VAL)	Unemployment (UC_VAL)	Workers Comp (WC_VAL)	Veterans (VET_VAL)	Retirement 1st source (RET_VAL1)
1987-1992	99,999	99,999	99,999	99,999	99,999	99,999	29,999	99,999
1993-1994	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999
1995-1997	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999
1998-2001	15,000	25,000	50,000	15,000	99,999	99,999	99,999	45,000
2002-2007	15,000	40,000	45,000	15,000	99,999	99,999	99,999	45,000

Year	Retirement 2nd Source (RET_VAL2)	Survivors 1st Source (SUR_VAL1)	Survivors 2nd Source (SUR_VAL2)	Disability 1st Source (DIS_VAL1)	Disability 2nd Source (DIS_VAL2)	Education Assistance (ED_VAL)	Financial Assistance (FIN_VAL)	Other (OI_VAL)
1987-1992	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999
1993-1994	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999
1995-1997	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999
1998-2001	45,000	50,000	50,000	35,000	35,000	20,000	30,000	25,000
2002-2007	45,000	50,000	50,000	35,000	35,000	20,000	30,000	25,000

Source: Current Population Survey Annual Demographic File Technical Documentation (1988-2002), Current Population Survey Annual Social and Economic Supplement Technical Documentation (2003-2008)

Appendix Table 1.4: Decomposition of GE(0) household income inequality into within-group and between-group inequality using alternative topcode correction methods.

Panel 1: Decomposition of GE(0) household income inequality by gender

	No Cell Mean				Consistent Topcode				Cell Mean				Internal			
Income Year	Within Female	Within Male	Total Within	Total Between	Within Female	Within Male	Total Within	Total Between	Within Female	Within Male	Total Within	Total Between	Within Female	Within Male	Total Within	Total Between
1975	0.290	0.265	0.278	0.001	0.267	0.265	0.278	0.001	0.297	0.273	0.285	0.001	0.298	0.273	0.286	0.001
1982	0.348	0.330	0.339	0.001	0.305	0.330	0.339	0.001	0.353	0.335	0.344	0.001	0.353	0.336	0.345	0.001
1992	0.356	0.324	0.340	0.001	0.330	0.324	0.340	0.001	0.372	0.341	0.357	0.001	0.373	0.342	0.358	0.001
1993	0.376	0.340	0.358	0.001	0.349	0.340	0.358	0.001	0.414	0.380	0.397	0.001	0.420	0.384	0.403	0.001
2004	0.418	0.386	0.402	0.001	0.399	0.386	0.402	0.001	0.457	0.426	0.442	0.001	0.467	0.445	0.456	0.001
2007	0.419	0.376	0.398	0.001	0.406	0.376	0.398	0.001	0.462	0.419	0.441	0.001	0.465	0.420	0.443	0.000

Panel 2: Decomposition of GE(0) household income inequality by race

	No Cell Mean						Consistent Topcode					
Income Year	Within Black	Within White	Within Hispanic	Within Other	Total Within	Total Between	Within Black	Within White	Within Hispanic	Within Other	Total Within	Total Between
1975	0.317	0.254	0.289	0.306	0.264	0.015	0.327	0.222	0.277	0.312	0.238	0.013
1982	0.410	0.303	0.373	0.378	0.322	0.018	0.403	0.252	0.357	0.353	0.279	0.016
1992	0.483	0.283	0.374	0.411	0.321	0.020	0.477	0.255	0.363	0.384	0.297	0.019
1993	0.509	0.298	0.371	0.445	0.337	0.022	0.500	0.269	0.359	0.416	0.312	0.020
2004	0.562	0.335	0.407	0.472	0.382	0.021	0.555	0.315	0.401	0.455	0.365	0.020
2007	0.569	0.326	0.416	0.427	0.377	0.021	0.565	0.312	0.409	0.409	0.365	0.020

	Cell Mean						Internal					
Income Year	Within Black	Within White	Within Hispanic	Within Other	Total Within	Total Between	Within Black	Within White	Within Hispanic	Within Other	Total Within	Total Between
1975	0.317	0.262	0.292	0.310	0.271	0.015	0.317	0.263	0.292	0.307	0.271	0.015
1982	0.411	0.308	0.377	0.384	0.327	0.018	0.411	0.309	0.377	0.384	0.327	0.018
1992	0.487	0.301	0.381	0.429	0.336	0.022	0.487	0.302	0.381	0.434	0.337	0.022
1993	0.532	0.339	0.392	0.487	0.374	0.024	0.534	0.345	0.393	0.500	0.379	0.025
2004	0.584	0.377	0.432	0.512	0.419	0.024	0.621	0.391	0.457	0.477	0.433	0.023
2007	0.602	0.371	0.438	0.468	0.417	0.025	0.607	0.375	0.438	0.479	0.419	0.024

Appendix Table 1.4 (continued)

Panel 3: Decomposition of GE(0) household income inequality by disability status[1]																
	No Cell Mean				Consistent Topcode				Cell Mean				Internal			
Income Year	Within Disabled	Within Not Disabled	Total Within	Total Between	Within Disabled	Within Not Disabled	Total Within	Total Between	Within Disabled	Within Not Disabled	Total Within	Total Between	Within Disabled	Within Not Disabled	Total Within	Total Between
1982	0.359	0.311	0.316	0.007	0.316	0.266	0.271	0.007	0.362	0.316	0.320	0.007				
1992	0.379	0.307	0.314	0.009	0.341	0.280	0.286	0.009	0.383	0.323	0.329	0.009	0.382	0.324	0.330	0.009
1993	0.397	0.323	0.331	0.011	0.362	0.295	0.302	0.011	0.405	0.361	0.366	0.013	0.404	0.365	0.369	0.013
2004	0.445	0.371	0.378	0.011	0.429	0.352	0.360	0.011	0.469	0.409	0.415	0.012	0.483	0.423	0.429	0.011
2007	0.476	0.362	0.373	0.011	0.468	0.350	0.361	0.010	0.505	0.404	0.413	0.011	0.510	0.406	0.416	0.011

Source: Author's calculations using Internal and Public Use March CPS data.

[1]Disability status is not available prior to 1980. Comparisons of household income for people with and without a disability excludes individuals age 15 and under, for whom disability status is not captured in the March CPS

Appendix Table 1.5: Decomposition of GE(2) household income inequality into within-group and between-group inequality using alternative topcode correction methods.

Panel 1: Decomposition of GE(2) household income inequality by gender

	No Cell Mean				Consistent Topcode				Cell Mean				Internal			
Income Year	Within Female	Within Male	Total Within	Total Between	Within Female	Within Male	Total Within	Total Between	Within Female	Within Male	Total Within	Total Between	Within Female	Within Male	Total Within	Total Between
1975	0.246	0.230	0.238	0.001	0.202	0.230	0.238	0.001	0.276	0.260	0.268	0.001	0.286	0.268	0.277	0.001
1982	0.281	0.261	0.271	0.001	0.227	0.261	0.271	0.001	0.297	0.277	0.287	0.001	0.301	0.279	0.290	0.001
1992	0.303	0.279	0.291	0.001	0.259	0.279	0.291	0.001	0.352	0.328	0.340	0.001	0.371	0.349	0.360	0.001
1993	0.314	0.285	0.299	0.001	0.272	0.285	0.299	0.001	0.446	0.412	0.429	0.001	0.648	0.592	0.619	0.001
2004	0.329	0.304	0.316	0.001	0.300	0.304	0.316	0.001	0.493	0.457	0.475	0.001	0.647	0.598	0.622	0.001
2007	0.313	0.289	0.301	0.001	0.286	0.289	0.301	0.001	0.473	0.440	0.456	0.001	0.578	0.524	0.550	0.000

Panel 2: Decomposition of GE(2) household income inequality by race

	No Cell Mean						Consistent Topcode					
Income Year	Within Black	Within White	Within Hispanic	Within Other	Total Within	Total Between	Within Black	Within White	Within Hispanic	Within Other	Total Within	Total Between
1975	0.274	0.217	0.262	0.267	0.227	0.012	0.274	0.173	0.229	0.200	0.183	0.011
1982	0.334	0.244	0.319	0.265	0.258	0.014	0.312	0.193	0.283	0.224	0.207	0.013
1992	0.397	0.254	0.359	0.283	0.275	0.016	0.373	0.211	0.321	0.250	0.232	0.015
1993	0.411	0.259	0.364	0.288	0.282	0.018	0.388	0.219	0.328	0.257	0.242	0.017
2004	0.376	0.275	0.352	0.320	0.299	0.018	0.357	0.247	0.337	0.290	0.271	0.017
2007	0.359	0.259	0.334	0.292	0.283	0.018	0.343	0.234	0.313	0.267	0.257	0.017

	Cell Mean						Internal					
Income Year	Within Black	Within White	Within Hispanic	Within Other	Total Within	Total Between	Within Black	Within White	Within Hispanic	Within Other	Total Within	Total Between
1975	0.274	0.248	0.284	0.282	0.257	0.012	0.275	0.258	0.286	0.273	0.266	0.012
1982	0.342	0.259	0.334	0.283	0.273	0.014	0.344	0.262	0.336	0.287	0.276	0.014
1992	0.417	0.302	0.392	0.331	0.323	0.018	0.422	0.323	0.392	0.354	0.343	0.018
1993	0.554	0.380	0.485	0.427	0.410	0.020	0.650	0.568	0.634	0.654	0.600	0.020
2004	0.524	0.426	0.485	0.475	0.455	0.020	0.562	0.577	0.661	0.555	0.603	0.019
2007	0.594	0.401	0.461	0.432	0.436	0.021	0.658	0.496	0.537	0.531	0.531	0.020

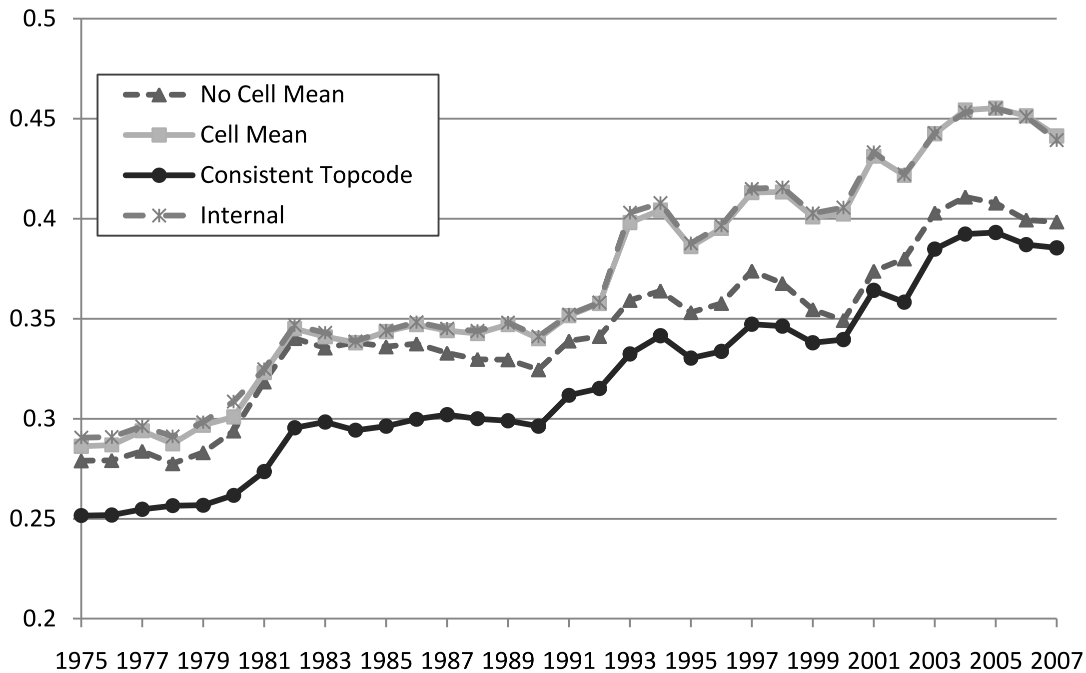
Appendix Table 1.5 (continued)

Panel 3: Decomposition of GE(0) household income inequality by disability status[1]

	No Cell Mean				Consistent Topcode				Cell Mean				Internal			
Income Year	Within Disabled	Within Not Disabled	Total Within	Total Between	Within Disabled	Within Not Disabled	Total Within	Total Between	Within Disabled	Within Not Disabled	Total Within	Total Between	Within Disabled	Within Not Disabled	Total Within	Total Between
1982	0.365	0.250	0.258	0.005	0.283	0.199	0.205	0.005	0.378	0.265	0.273	0.005				
1992	0.399	0.267	0.276	0.007	0.312	0.223	0.230	0.007	0.424	0.313	0.322	0.007	0.420	0.333	0.341	0.007
1993	0.387	0.271	0.280	0.008	0.308	0.230	0.237	0.008	0.435	0.396	0.405	0.010	0.454	0.565	0.571	0.010
2004	0.380	0.291	0.299	0.008	0.343	0.264	0.271	0.008	0.513	0.440	0.451	0.009	0.763	0.576	0.593	0.009
2007	0.374	0.276	0.284	0.008	0.353	0.251	0.259	0.008	0.513	0.419	0.429	0.008	0.654	0.508	0.522	0.008

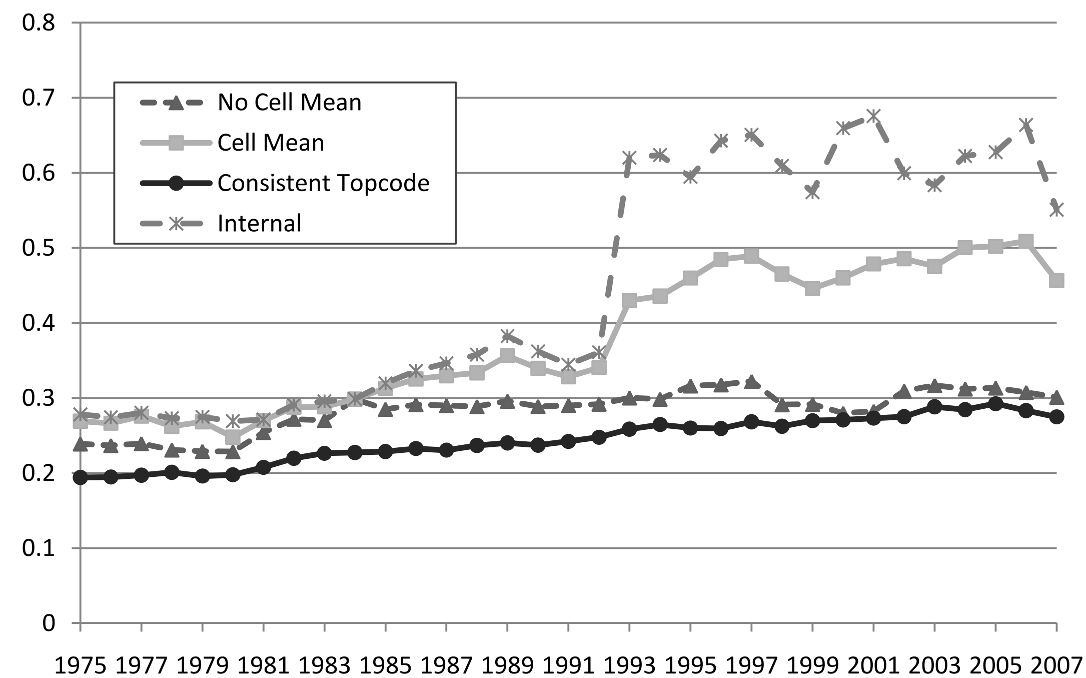
Source: Author's calculations using Internal and Public Use March CPS data

[1]Disability status is not available prior to 1980. Comparisons of household income for people with and without a disability excludes individuals age 15 and under, for whom disability status is not captured in the March CPS



Source: Author's calculations using Internal and Public Use March CPS data

Appendix Figure 1.1: GE(0) household income inequality index using alternative topcode correction methods



Source: Author's calculations using Internal and Public Use March CPS data

Appendix Figure 1.2: GE(2) household income inequality index using alternative topcode correction methods

REFERENCES

Acemoglu, Daron and Joshua Angrist. 2001. "Consequences of Employment Protection? The Case of the Americans with Disabilities Act." *Journal of Political Economy* 109(5): 915-957.

Atkinson, Anthony B., Lee Rainwater, and Timothy Smeeding. 1995. *Income Distribution in OECD Countries. Evidence from the Luxembourg Income Study*. Social Policy Studies No. 18. Paris: Organization for Economic Cooperation and Development.

Blau, Francine D. and Lawrence M. Kahn. 1997. "Swimming Upstream: Trends in the Gender Wage Differential in the 1980s." *Journal of Labor Economics* 15(1): 1-42

Blau, Francine D. and Lawrence M. Kahn. 2000. "Gender Differences in Pay." *Journal of Economic Perspectives* 14(4): 75-99.

Burkhauser, Richard V., J.S. Butler, Shuaizhang Feng, and Andrew J. Houtenville. 2004. "Long Term Trends in Earnings Inequality: What the CPS Can Tell Us." *Economics Letters* 82(2): 295-299.

Burkhauser, Richard V., Kenneth A. Couch, Andrew J. Houtenville, and Ludmila Rovba. 2004. "Income Inequality in the 1990s: Re-Forging a Lost Relationship?" *Journal of Income Distribution* 12(3/4): 8-35.

Burkhauser, Richard V., Shuaizhang Feng, and Stephen P. Jenkins. 2009. "Using a P90/P10 ratio to measure inequality trends with the public use Current Population Survey: A view from inside the Census Bureau Vaults." *The Review of Income and Wealth* 55(1): 166-185.

Burkhauser, Richard V., Shuaizhang Feng, Stephen P. Jenkins, and Jeff Larrimore. Forthcoming. "Trends in United States Income Inequality Using the March Current Population Survey: The Importance of Controlling for Censoring." *Journal of Economic Inequality.*

Burkhauser, Richard V., Shuaizhang Feng, and Jeff Larrimore. Forthcoming. "Measuring Labor Earnings Inequality Using Public use March Current Population Survey Data: The Value of Including Variance and Cell Means when Imputing Topcoded Values" *Economic Letters.*

Card, David and John E. DiNardo. 2002. "Skill-Biased Technological Change and Rising Wage Inequality: Some Problems and Puzzles." *Journal of Labor Economics* 20(4): 733-783.

Couch, Kenneth and Mary C. Daly. 2002. "Black-white wage inequality in the 1990s: A decade of progress." *Economic Inquiry* 40(1): 31-41.

Cowell, Frank A, 2000. "Measurement of Inequality" in: Anthony B. Atkinson and François Bourguignon (eds), *Handbook of Income Distribution*, pp. 87–166. North Holland, Amsterdam: Elsevier.

Daly, Mary C. and Robert G. Valletta. 2006. "Inequality and Poverty in the United States: The Effects of Rising Dispersion of Men's Earnings and Changing Family Behavior." *Economica* 73(289): 75-98.

Denavas-Walt, Carmen, Bernadette D. Proctor, and Jessica C. Smith, U.S. Census Bureau. 2009. *Income, Poverty, and Health Insurance Coverage in the United States: 2008*. (August), Current Population Reports P60-235. Washington DC: GPO.

Devine. Theresa J. 1994. "Characteristics of Self-Employed Women in the United States." *Monthly Labor Review* 117(3): 20-34.

Feng, Shuaizhang, Richard V. Burkhauser, and J.S. Butler. 2006. "Levels and Long-Term Trends in Earnings Inequality: Overcoming Current Population Survey Censoring Problems Using the GB2 Distribution." *Journal of Business and Economic Statistics* 24(1): 57–62.

Gottschalk, Peter and Sheldon Danziger. 2005. "Inequality of Wage Rates, Earnings and Family Income in the United States, 1975-2002." *Review of Income and Wealth,* 51(2): 231–254.

Hotchkiss, Julie L. 2004. "A Closer Look at the Employment Impact of the Americans with Disabilities Act" *Journal of Human Resources* 39(4): 887-911.

Jenkins, Stephen P., Richard V. Burkhauser, Shuaizhang Feng, and Jeff Larrimore. Forthcoming. "Measuring Inequality using Censored Data: A Multiple Imputation Approach to Estimation and Inference." *Journal of the Royal Statistical Society: Series A*.

Jenkins, Stephen P. and Van Kerm, P. 2009. "The Measurement of Economic Inequality" in: Wiemer Salverda, Brian Nolan and Timothy M. Smeeding (eds), *The Oxford Handbook on Economic Inequality*. New York: Oxford University Press.

Jones, Arthur F. and Daniel H. Weinberg, U.S. Census Bureau. 2000. Current Population Reports P60-204. *The Changing Shape of the Nation's Income Distribution*. Washington, DC: GPO.

Juhn, Chinhui. 2003. "Labor Market Dropouts and Trends in the Wages of Black and White Men." *Industrial and Labor Relations Review* 56(4): 643-662.

Juhn, Chinhui, Kevin M. Murphy, and Brooks Pierce. 1991. "Accounting for the Slowdown in Black-White Wage Convergence." in Marvin Kosters (Ed), *Workers and their Wages*. Washington, DC: AEI Press.

Karoly, Lynn A. and Gary Burtless. 1995. "Demographic Change, Rising Earnings Inequality, and the Distribution of Personal Well Being, 1959-1989." *Demography* 32(3): 379-405.

Larrimore, Jeff, Richard V. Burkhauser, Shuaizhang Feng and Laura Zayatz. 2008. "Consistent Cell Means for Topcoded Incomes in the Public Use March CPS (1975-2007)." *Journal of Economic and Social Measurement* 33(2/3): 89-128.

Levy, Frank and Richard J. Murnane. 1992. “U.S. Earnings Levels and Earnings Inequality: A Review of Recent Trends and Proposed Explanations.” *Journal of Economic Literature,* 30(3): 1333–1381.

Ryscavage, Paul. 1995. “A Surge in Growing Income Inequality?” *Monthly Labor Review* 118(8): 51-61.

Welniak, Edward J. 2003. “Measuring Household Income Inequality Using the CPS.” in James Dalton and Beth Kilss (Eds.), *Special Studies in Federal Tax Statistics 2003*. Washington, DC: Statistics of Income Directorate, Internal Revenue Service.

CHAPTER 2

THE IMPACT OF CHANGING EARNINGS DISTRIBUTIONS AND HOUSEHOLD CHARACTERISTICS ON US INCOME INEQUALITY TRENDS SINCE 1967

Abstract

While much of the research on the rise in inequality in the United States since the 1960s has focused on labor earnings inequality, there is little evidence regarding how closely these labor earnings inequality trends correlate to the broader measure of household income inequality. This paper first compares male and female labor earnings inequality to that of household income. It then uses a shift-share analysis to analyze the change in income inequality accounted for by changes in male and female labor earnings distributions and changing household characteristics. In doing so, it is evident that the factors contributing to the rapid rise in household income inequality in the 1970s and 1980s differ substantially from those contributing to the slower increase in the 1990s. In contrast to findings for the 1970s and 1980s, in more recent years increases in male earnings inequality largely account for the changes in household income inequality while declines in the correlation between spouses' earnings have mitigated household income inequality growth.

2.1 Introduction

Since researchers first observed a dramatic rise in labor earnings and income inequality in the early 1980s, there has been a strong interest in understanding these trends. Much of this research has focused on labor earnings inequality among full-time workers and has been motivated by concerns regarding how labor is compensated in the labor market. Among other factors, researchers have explored how the relative wages of high and low skill workers changed due to increasing returns to post-secondary education (Lemieux 2006a), skill-biased technological change (Juhn, Murphy and Pierce 1993, Berman, Bound and Griliches 1994, Autor, Katz and Kearney 2008), the decline in the real minimum wage (Dinardo, Fortin, and Lemieux 1996, Lee 1999, Card and Dinardo 2002, Lemieux 2006b), and the decline in unionization (Dinardo, Fortin, and Lemieux 1996).

A smaller literature has considered the related question of how these changes in labor earnings inequality relate to changes in household income inequality.[18] The most naïve researchers believe that household income inequality and labor earnings inequality are equivalent. This would be true if all households contained only one worker, that worker worked full-time, and there were no non-labor earnings or government transfers. However, as illustrated by Stigler's (1946) research on minimum wage legislation, when there is more than one worker per household, when workers work different hours, or when there are other sources of income in a household, then all these factors must be considered in evaluating the impact of changes in earnings inequality on trends in income inequality.

Using a shift-share analysis, this paper explores this relationship and the extent to which changes in male and female labor earnings inequality have translated into changes in household income inequality. Additionally, it attempts to better estimate how shifts in employment rates, hours worked, marriage rates, and the correlation between spouses' earnings, have influence household income inequality trends since 1967.

[18] The terms earnings and labor earnings are used interchangeably in this paper to refer to earnings from wages and salaries, self-employment, or farm-employment. The terms income and household income are used interchangeably to refer to all income within a household from any income source.

There is a small literature that has previously explored elements of this relationship between U.S. labor earnings inequality and household income inequality. Lerman and Yitzhaki (1985), Karoly and Burtless (1995), Cancian and Reed (1998), and Bayez and Couch (2008) decompose the Gini coefficient for family income into the contributions from component income sources using Fei, Ranis, and Kuo's (1978) Gini decomposition. Burtless (1999) and Daly and Valetta (2006) use a shift-share analysis, comparing family income inequality across two years and observing how inequality changes would have differed had only certain income components changed over that time. In addition to the literature considering these questions in the United States, there is a somewhat more expansive literature using similar decomposition and shift-share methods considering the relationships between labor earnings inequality and household income inequality in an international context (see, for example, Mookherjee and Shorrocks 1982, Jenkins 1995, and Jenkins 1996 for the UK, Fournier 2001 for Taiwan, Del Boca and Pasqua 2003 for Italy, and Pasqua 2008 for Europe).

Such studies report mixed findings on the strength of the relationship between male earnings inequality and household income inequality. For example, Karoly and Burtless (1995) suggest that rising correlations between spouses' earnings is as important as rising male earnings inequality in explaining rising family income inequality. Burtless (1999) also finds that a minority of the rise in income inequality in the 1980s was due to male earnings inequality, with changes in marriage patterns and an increase in single-headed families contributing substantially to the increase in income inequality. In a later paper, Burtless (2009) notes that marriage patterns continue to be important although he also observes that they cannot explain periods of rapid inequality growth because changes in marriage patterns occur slowly over time. Daly and Valetta (2006), on the other hand, place a higher importance on male earnings inequality changes, suggesting that they explain the majority of the rise in income inequality since 1979.

In contrast to the work by Daly and Valetta (2006), Burtless (1999), and Burtless and Karoly (1995) which evaluate inequality trends by looking at a single year per business cycle, Gottschalk and Danziger (2005) compare annual male earnings inequality trends to those for income inequality. Using P90/P10 ratios, Gottschalk and Danziger observe that while the overall rise in male earnings inequality since 1975 is similar to the overall rise in income inequality, the timing of these increases are different. They therefore suggest that other factors are likely contributing to the observed income inequality increases. Their findings illustrate the importance of considering complete trends in inequality, rather than concentrating only on single years. Had Gottschalk and Danziger only considered the beginning and ending years of their sample period, the rise in income inequality and male earnings inequality would have seemed quite similar. By comparing the complete trends they reached a different conclusion.

One limitation of the work by Gottschalk and Danziger (2005) and other inequality researchers, however, is that they do not have access to the complete income distribution because the Census Bureau censors top incomes in the March CPS data. This topcoding makes it difficult to consistently observe changes at the top of the distribution. A common approach used by Gottschalk and Danziger to limit the influence of topcoding on inequality results is to measure inequality using the P90/P10 ratio. Since the P90/P10 ratio is unaffected by changes to the distribution above the 90^{th} percentile, the distortionary effects of topcoding on P90/P10 ratios are limited when compared to most other inequality measures.[19]

[19] Although the P90/P10 ratio does reduce the impact of topcoding, Burkhauser, Feng, and Jenkins (2009) show that P90/P10 ratios do not completely overcome the problem because CPS topcoding is performed on each sources of income separately. Nevertheless, the effects of topcoding are smaller than

However, the features of the P90/P10 ratio that make it beneficial for limiting the distortionary effects of topcoding also make it an imperfect measure of inequality when compared to other inequality indices. As described by Jenkins and Van Kerm (2009), there are four key properties that are desirable in an inequality index: scale invariance, replication invariance, symmetry (or anonymity), and satisfying the Pigou-Dalton transfer principle.[20] While the P90/P10 ratio satisfies the first of these three principles, it does not satisfy the Pigou-Dalton transfer principle since a transfer of income will only impact the P90/P10 ratio if the transfer impacts the income of the individual at the 90^{th} or 10^{th} percentile. This is in contrast to other commonly used measures of inequality such as the Gini coefficient, the Generalized Entropy family of inequality indices including the Thiel index, and the Atkinson indices. Each of these inequality measures satisfy all four desirable properties for measuring inequality.

This distinction would be relatively innocuous if it did not influence inequality trends. However, this is not the case. Jenkins and Van Kerm (2009) illustrate that cross-country rankings of income inequality are affected by the choice between a P90/P10 ratio and a Gini index. Furthermore, Burkhauser, Feng, and Jenkins (2009) show that even after correcting for topcoding the inequality trends using P90/P10 ratios are quite different from those found using the Gini coefficient – and this difference varies for labor earnings and household income. Therefore, using the P90/P10 ratio to analyze the contribution of male earnings inequality to household income inequality may lead to different results than those found using Gini

for Gini coefficients or other measures of inequality that incorporate dispersion through the entire distribution.

[20] Scale invariance states that the index is insensitive to a rescaling of the unit of income measurement, such as shifting from measuring income in dollars to measuring it in cents. Replication invariance states that the index is insensitive to a replication of all individuals and their incomes, thus making the index comparable for populations of different sizes. The symmetry axiom states that only the income of individuals, and not personal characteristics, impact the measure of inequality. Finally, the Pigou-Dalton transfer principle states that a small transfer of income from an arbitrarily chosen person to a person lower in the distribution, while keeping the transferor richer overall, reduces income inequality. For a further discussion of these properties see Jenkins and Van Kerm (2009).

coefficients.

With access to the internal CPS data, I am able to lift these topcoding constraints and observe incomes in the distribution above those in the public use CPS. As a result, this paper examines the relationship between labor earnings and income inequality across the entire distribution using broad-based inequality metrics that satisfy all four desirable properties of inequality indices. Since it is one of the most commonly used measures of income inequality, this paper focuses on the Gini coefficient when analyzing trends in income inequality.[21]

Additionally, to further understand the relationship between changes in male earnings inequality and changes in household income inequality, the analysis in this paper goes beyond the comparison of trends performed by Gottschalk and Danziger and examines the extent to which increases in household income inequality are attributable to changes in each income source and changes in household characteristics. This additional analysis is performed using a shift-share procedure similar to that used by Burtless (1999) and Daly and Valetta (2005).

This paper further adds to the analysis by considering annual trends in inequality rather than just one year per business cycle, examining the entire income distribution rather than just the portion under the topcoding threshold, separately evaluating different types of changes occurring within income sources, and expanding the studied time-frame. Using this analysis, the results reconcile the varied results on the importance of each income component on household income found in the literature. They also show that different factors account for household income inequality changes in the 1990s than in earlier decades, indicating that the relationship between earnings inequality and household income inequality has evolved over time.

[21] Similar results can be obtained using the GE*(1)* Generalized Entropy measure (the Theil Index) which, like the Gini coefficient, is relatively sensitive to changes at the middle of the income distribution. Results based upon the GE*(1)* inequality index are available in the Appendix.

The results in this paper provide new information on the factors accountable for rises in income inequality, although this accounting should not be viewed in a causal sense due to the complex behavioral interactions occurring across the factors considered. For example, if increases in female wage rates empowered more women to live independently this could lead to a decline in marriage rates. Thus, from a causal perspective inequality changes accounted for by changing marriage rates may actually be caused by changes in female employment possibilities. Similarly, if increases in public transfers induce individuals to leave the labor market, then the public transfers – which in this strict accounting sense reduce inequality – could have behavioral implications that increase inequality through other channels. Nevertheless, given the relatively little research exploring factors contributing to household income inequality changes, it is valuable to first consider such an accounting approach to changes in income inequality which can then direct future research into the impact of these factors in a causal sense.

2.2 Data

Topcoding in the March CPS data. This analysis derives from access to internal CPS data, which is identical to the data used by the Census Bureau for producing their official income statistics (Denavas-Walt, Proctor, and Smith 2009). These data measure top incomes much better than the data released in public use CPS files. To protect the confidentiality of its respondents, the Census Bureau censors ("topcodes") each of the income sources received by individuals in the public use data and the extent of topcoding varies over time. As a result, the public use CPS data traditionally allowed researchers to at-best consistently measures inequality for the 95 percent of the income distribution below the topcode thresholds and at-worst provide inconsistent estimates of inequality due to variations in topcoding (See Feng, Burkhauser, and

Butler 2006 and Larrimore et al. 2008 for comparisons of inequality results using various topcode correction procedures).

The internal CPS data does not have the same topcode constraints. While some censoring occurs in the internal CPS data, this censoring is much less extensive than that seen in the public use CPS data. Less than 1 percent of the population has their household income censored in the internal data in any given year, while several recent years have upwards of 5 percent of individuals with household income topcoded in the public use data. The limited internal censoring exists mainly to minimize the impact of recording errors and prevent volatility in annual statistics (Semega and Welniak, 2007). Furthermore, since the internal data is the same as that used by the Census Bureau for their official income statistics, this censoring is no more restrictive than that which is incorporated into the government's official inequality statistics.

One limitation when calculating long-term trends that cannot be corrected by using internal data is the potential for survey design changes to influence the results. While the March CPS data is largely consistent over time, there were substantial changes between 1992 and 1993 when the Census Bureau implemented computerized data collection along with several other data collection procedure changes (See Ryscavage, 1995 and Jones and Weinberg, 2000 for further discussion of these changes). These changes improved the Census Bureau's accuracy in recording incomes, particularly at the top of the distribution. However, this also led to a large artificial increase in inequality so the Census Bureau recommends against making comparisons over these years. Thus, 1992-1993 are separated in the results due to these data comparability problems.

Defining Income. The results in this paper focus on the size-adjusted household income of persons, including both labor and non-labor earnings. This income measure is commonly used in US and cross-national studies of income inequality (see, for example, Atkinson, Rainwater and Smeeding 1995, Gottschalk and Smeeding 1997, Atkinson and Brandolini 2001, and Burkhauser et al. Forthcoming). It assumes that income is shared equally among all household members, so each individual in the household receives the same income. To account for economies of scale in household consumption, household income is divided by the square root of household size to obtain size-adjusted household income.[22] This aggregation and size adjustment is performed at the level of the income source. For example, all individuals in a household are assigned the same male-head labor earnings, which is equal to the earnings received by the male household head divided by the square root of household size.[23] Such a procedure is necessary to ensure that all income is accounted for when considering the impact of earnings source changes on income inequality.

As is common in the income inequality literature, individuals in group quarters or in a household containing a member of the military are excluded. Additionally, unlike in labor earnings analyses where the sample population is often restricted to working age individuals, analyses of household income inequality generally include all individuals regardless of age. When analyzing household income inequality, this paper will do the same.

2.3 Comparing Gini coefficient trends for labor earnings and household income

There are several ways to examine the relationship between earnings inequality and

[22] Dividing by the square-root of the household size is the most commonly used case of the economies of scale size-adjustments proposed by Buhmann et al. (1988) where size-adjusted HH income = (total HH income) / (HH size)$^{\alpha}$, with α=1 implying no economies of scale and α=0 implying infinite economies of scale. Setting α=0.5 closely matches the adjustments for household size implied by the Census Bureau poverty thresholds (Ruggles 1990).

[23] The household head refers to the Census householder in years since 1980 and the Census household head prior to that time. The definition of the household head is the person (or people) in whose name the housing unit is owned or rented. In cases where there is no such person, it may refer to any adult member of the household excluding boarders (US Census Bureau, 2008). In cases of married individuals, while the Census arbitrarily considers one person to be the householder, this paper refers to both as being household heads.

household income inequality. One way is to simply compare the inequality levels and trends for household income to those for labor earnings. This is the procedure used by Gottschalk and Danziger (2005) to explore inequality trends using P90/P10 ratios and similar comparisons are made here using Gini coefficients.

The comparison used here starts with a definition of labor earnings commonly used in the earnings inequality literature – personal, non-size-adjusted labor earnings among working age individuals who have positive earnings (Card and DiNardo 2002 and Gottschalk and Danziger 2005 use a similar definition, although restricting earnings to wage earnings rather than the more inclusive labor earnings definition). Working age individuals are defined in this paper as those aged 22-62. This sample is divided by gender, reflecting the fact that the labor earnings distributions differ for men and women. The Gini coefficients using this income definition by gender are provided in Columns 1 and 3 of Table 2.1 for the trough years of each business cycle since 1967.[24] While there are valuable insights to be gained from more carefully analyzing the annual trends in inequality, which will be discussed in more detail in the sections below, such a comparison across trough years allows for a snapshot of inequality trends devoid of cyclical business-cycle variations.

Both male and female earnings inequality among working individuals has risen since 1975, although it increased more rapidly for men. However, while these series are informative for understanding inequality in labor market compensation, they do not necessarily reflect inequality in society as a whole. For example, both series exclude people with zero earnings. Including these individuals, but still analyzing

[24] Trough years of business cycles are defined here based on troughs in income which generally lag macroeconomic growth. While not trough years, 1967 and 2007 are included in Table 1, and all tables of business cycle trough years, because they are the first and last years of data available. 1993 is also included along with the actual trough year of the 1990s business cycle, 1992, due to the Census redesign in 1993 that limits data comparability from 1992-1993. Thus, the inclusion of 1993 in the tables separates out any inequality changes due to the survey methods from actual changes in inequality that occurred during the 1980s and 1990s business cycles.

Table 2.1: Gini coefficients for male and female labor earnings and size-adjusted household income in business-cycle trough years since 1967.

	Male Labor Earnings		Female Labor Earnings		Size-Adjusted Household Income	
	Working-age with earnings	All Working-age	Working-age with earnings	All Working-age	All Working-age	All ages
1967	0.328	0.358	0.433	0.690	0.343	0.363
1975	0.352	0.408	0.416	0.652	0.338	0.359
1983	0.385	0.454	0.416	0.604	0.366	0.386
1992	0.414	0.478	0.418	0.567	0.383	0.404
1993	0.444	0.510	0.437	0.579	0.408	0.427
2004	0.447	0.520	0.427	0.574	0.417	0.434
2007	0.439	0.511	0.425	0.568	0.412	0.431

Source: Authors calculations using Internal March CPS data (1967-2007)
Notes: While not trough years, 1967 and 2007 are included in tables of trough years because they are the first and last years of data available. 1993 is also included to separate the large artificial increase in inequality that occurred between 1992-1993 due to changes in the March CPS data collection procedures from actual changes occurring before and after that time.

personal, non-size-adjusted labor earnings among working age individuals (Columns 2 and 4 of Table 2.1) leads to higher levels of inequality for both sexes. Additionally, the choice to include or exclude individuals with no earnings impacts not just the levels of inequality but also their trends. This is most evident when comparing trends between the two female labor earnings series. Among working age women who work, labor earnings inequality increased slightly since 1975, while labor earnings inequality among all working age women declined dramatically as a result of increases in female employment rates.

Including individuals with no labor earnings is just one way that household income inequality trends could differ from those for labor earnings. Household income includes not just own-labor earnings, but also non-labor income such as public transfers, interest, dividend, and rental income. The inclusion of these additional income sources and assuming sharing of income across members of the household will further change levels and trends of inequality. This can be seen in Column 5 of Table

2.1 which considers inequality of size-adjusted household income of individuals but still restricts the sample only those of working age. Including these additional factors generates lower levels of inequality than that seen for either male or female labor earnings. Additionally, the growth in household income inequality for working age individuals is moderately slower over the 40 year period than that seen for labor earnings inequality among working age men alone.

A final distinction in understanding the differences between household income and labor earnings inequality is the age-range of analysis. Researchers interested in household income inequality are interested in inequality across all age ranges including children and the aged, rather than just among working age adults. Since children typically live with their parents and therefore are assumed to share the parents income for consumption, the inclusion of children mainly impact inequality only by increasing the importance of large households in the analysis. The income composition of the aged, however, is quite different from that of working age individuals. As can be seen by comparing Column 5 of Table 2.1, which considered size-adjusted household income for working age individuals, to Column 6 of Table 2.1, which expands the sample to include individuals of all ages, including the aged and children increases the levels of household income inequality when compared to that of working age individuals.

A casual observer may compare column 1 – the labor-earnings inequality for working age men who work – to Column 6 of Table 2.1 – the size-adjusted household income inequality for the entire US population and observe that since the levels are similar they must be measuring approximately the same things. However, as should be apparent through the discussion above, the underlying factors that can influence household income inequality extend well beyond just male labor earnings. Thus, the observation that levels of inequality are so similar for these two series is therefore largely coincidental.

Of course, while it is not a one-to-one relationship the changes in male labor earnings inequality can, and do, influence levels and trends in household income. To the extent that male earnings are an important component of household income, an increase in male earnings inequality may increase household income inequality. However, a basic comparison of inequality trends across series is limited in its ability to explain the extent to which a single income source accounts for the rise in household income inequality. This is because there are three ways that changes to an income source can influence household income inequality. The first is through changes to the level of inequality of the source. The second is through changes to the share of income coming from the source. And the third is through changes in the correlation between income from that source and income from other sources. Comparing the trends in source-level inequality to the trend in household income inequality only considers the first of these three contribution paths. Understanding the full impact of how male or female earnings changes account for household income inequality changes requires analyzing the other two pathways as well.

Additionally, comparing inequality trends cannot provide information about contributions to rising income inequality that are not due to changes in source-level income distributions but are instead due to changes in how individuals form households. Given these limitations, an alternate shift-share approach is used to further examine the income sources and household demographic shifts responsible for the increase in income inequality over the past 40 years.

2.4 Method of decomposing the increase in household income inequality

To decompose the change in household income inequality into that attributable to male earnings changes and to other income and demographic changes, the shift-share analysis starts with the March CPS sample from 1975.[25] Changes that could impact inequality are then added one at a time and the resulting increase in income inequality is compared to the increase that would have occurred had the specified factor remained unchanged. This yields the inequality changes that can be accounted for by each factor. In this procedure, potential causes are divided into three categories: changes to the prevalence of population groups, changes to the distribution of incomes from a given source, and changes to the correlations of income across income sources. Each of these categories requires slightly different methods for capturing their relationships with household income inequality changes.

Changes in the prevalence of population groups. The portion of household income inequality changes attributable to changes in population group size is determined using a shift-share approach commonly used to separate changes into their factor components (Atkinson 1998, Burtless 1999**)**. This procedure starts with the population in time *t* where the income distribution is described by the income frequency density function, $\varphi^t(y)$. The population in time *t* can be divided into K mutually exclusive subgroups where N_k^t the fraction of the population belonging to group *k* in time *t* is ν_k^t. Each subgroup's income distribution is described by the income frequency density function $\varphi_k^t(y)$.

As noted by Jenkins (1996), the population income frequency, $\varphi^t(y)$, equals the weighted sum of the subgroup frequencies, $\varphi_k^t(y)$, with the weight equal to the population share of the subgroups, ν_k^t. As a result, changes to group size can impact the population income distribution and the inequality metrics calculated based off of it even if the subgroup income distributions remain unchanged. This is because if one

[25] 1975 was chosen as a base-year both because it is commonly used as the initial year in studies of long-term income inequality trends (see, for example, Gottschalk and Danzinger, 2005) and because a comparison of household-income and labor-earnings trends indicates that 1975 is the beginning of an era where the relationship between male-earnings inequality and household income inequality changed. Alternate base-years were tested and did not substantially impact the findings in this paper.

subgroup's population share increases over time then the population income distribution will increasingly approximate the growing group's income distribution as its weight increases in the population.

For this reason, changes in population group sizes have the potential to impact income inequality without changing the underlying income distributions of those groups. For example, if a subgroup with relatively high inequality increases in size then, holding the subgroup income inequalities constant, then its larger subgroup size it will increase overall income inequality.

The importance of the relative size of these population groups can be analyzed by considering how income inequality trends would have differed if the income distributions of individuals within each group remained constant and only the size of the groups had changed. Thus, suppose that ν_k^t percent of the population is in subgroup *k* in year *t* and $\nu_k^{t'}$ percent of the population is in the same subgroup in year *t'*. Then the impact of the change in the group size can be captured by reweighting observations from time *t* such that the fraction of the population in group *k* is $\nu_k^{t'}$. This is accomplished by replacing the observation weight for each individual *i* in group *k* in year *t*, $W_{i,k}^t$, with:

$$\widehat{W}_{i,k}^{t'} = W_{i,k}^t \left(\frac{\nu_k^{t'}}{\nu_k^t}\right) \tag{2.1}$$

This increases the weight of individuals in groups that are more prevalent in year t' than in year t and reduces the weight of individuals in groups that are less prevalent in year t' than in year t.

Changes in source-level income distributions within population groups. In addition to changes in the prevalence of each population group, there have also been changes to the distribution of incomes of individuals within each subgroup. These changes can come from any one of the income sources received from individuals, including male labor earnings, female labor earnings, public transfers, or non-labor income. This is the second factor considered: changes to the source-level distribution of incomes.

The portion of household income inequality changes attributable to changes in population group size is determined using a rank-preserving income exchange. The procedure used is similar to that used by Burtless (1999) and Daly and Valetta (2005), although there are several important differences. First, the income exchange is a conditional on full-time or part-time employment status. This is in contrast to earlier work that either performed an unconditional income exchange, or only conditioned on having any employment rather than on intensity of employment. This conditioning allows me to separate the impact of changes to work-intensity from changes in earnings inequality among individuals working full-time, part-time, or not working. Additionally, in this paper the exchange is performed in two steps – first allowing the inequality of source-level incomes to vary while holding the conditional mean earnings levels constant and then allowing both the source-level inequality and real earnings level to vary. This two-step process differentiates the impact of these two ways that source-level income distribution changes can influence income inequality.

In order to perform the conditional rank-preserving income exchange, individuals are ordered in each subgroup *k* from low to high based on their income from source *f* and assign them a source-level income rank, r_{ikf}^{t}, based on this ordering. Rank 1 represents the individual with the lowest income from source *f* among group *k* and rank N_k^t being the individual with the highest income from source *f* among group *k*. Note that an individual's rank will generally not be the same across two sources of income, so each individual has a separate rank for all income sources analyzed. The source-level income of the individual at any given rank in time *t* can therefore be denoted as y_{kfr}^{t} where $y_{kfr_1}^{t} \geq y_{kfr_2}^{t}$ for $r_1 > r_2$.

To understand how changes in the source-level distribution of incomes among individuals in group *k* relate to household income inequality, rank-correlations across income sources are assumed to remain unchanged. To observe the impact of changing the distribution of earnings from each source, starting with the income distribution of year *t*, each individual is assigned the income from source *f* of the individual in year *t'* with the same rank in the source-level distribution. Income from all other sources remains unchanged. This procedure preserves the conditional earnings rank of each individual but captures the change in the source-level income distribution among individuals of the group.

If there were the same number individuals in each group in all years, this process would be quite straightforward using a simple one-to-one replacement. Unfortunately, this is not the case. In general, $N_k^t \neq N_k^{t'}$, which means that individuals with the same numerical ranks are at different points in the income distribution in each year. Thus, an exact rank replacement would truncate the top of the distribution when subgroup membership expands. Therefore, to avoid this problem ranks in year *t* are rescaled to match the number of observations in year *t'*:

$$\hat{r}_{ikf}^t = r_{ikf}^t \left(\frac{N_k^{t'}}{N_k^t}\right) \qquad (2.2)$$

Since these new ranks are not restricted to integer values as necessary for replacement, they are rounded up or down randomly in proportion to the decimal value of the estimated rank. So, for example, if the rescaled rank of an individual is 100.2, he will be assigned a rank of 100 with probability of 0.8 and a rank of 101 with probability 0.2. Once the rescaled ranks from year *t* are determined, individuals are assigned the income from year *t'* corresponding to the rescaled rank, $y_{kf\hat{r}}^{t'}$. This income is then added to the individual's income from all other sources in year *t* to determine how income inequality would have changed had only income from source *f* changed and income from all other sources remained constant.

To separate the impact of changes to the source-level inequality from real earnings growth of the source, the rank-preserving income exchange described above is divided into two components: changes to the mean-preserving source-level income distribution and changes to the inflation-adjusted source mean incomes. First, the analysis is performed keeping the source-level mean incomes constant over time. This captures the impact of the change in dispersion of source-level income without capturing the change in real earnings levels. Second, the real earnings growth is included as well, capturing the additional impact on inequality of income from specific sources growing faster or slower than the rate of inflation over time.

Changes in income-source rank correlations within population groups. The previous two methods each assume that the rank correlation of income sources is unchanged over time. Thus, if the man at rank *n* in the conditional male earnings distribution is married to the woman at rank *m* in the conditional female earnings distribution in year *t*, then the procedure described above assumed that the man at rank *n* in the male earnings distribution in all future years is also married to the woman at rank *m* in the female earnings distribution. The third area of analysis removes this assumption and considers how changes to the rank correlation of earnings sources impact household income inequality.

The procedure for incorporating changes to the rank correlation is as follows, using similar procedures to those used by Burtless (1999) and Fournier (2001). Taking the rescaled income source ranks described above, $\hat{r}_{ikf}^{t}$, each individual's ranks are observed to establish rank pairings across the income sources. The relationship between changes to rank correlations and changes to household income inequality can be captured by replacing the rank pairings from year *t* with the rank pairings from year *t'* - and rearranging the source incomes to correspond to the new rank pairings. This revised source-level income is then added to the income from all other income sources to calculate the estimated Gini coefficient and observe the change in income inequality attributable to the correlation change.

Robustness of results to changes in the order of analysis. A known limitation of this type of shift-share analysis is that the results are sensitive to the order in which the component factors are analyzed (Jenkins 1995, Fournier 2001, Daly and Valetta 2006). One approach for analyzing the factors is to add in the changes to factor components sequentially (Daly and Valetta 2006). Thus, for example, one would first observe how household income inequality changes since year *t* if nothing changes except for marriage rates. Then, one would observe how household income inequality changes since year *t* if nothing changes except for marriage rates *and* the male earnings distribution. The portion of the total household income inequality change accounted for by each factor is therefore the additional change in inequality observed from the additional component – beyond the change accounted for in the previously analyzed factors. The advantage of this method is that it ensures that exactly 100 percent of the actual change in household income inequality is captured by the component factors. However, the disadvantage is that altering the order in which factors are analyzed has the potential to change the results.

An alternate method of performing the shift-share analysis is to always start with the income distribution in year *t* and only change a single factor (Burtless 1999). For example, using this approach one would observe how household income inequality differs from that seen in year *t* if nothing changes except for marriage rates. Then, one would observe how household income inequality differs from that seen in year *t* if nothing changes except for the male earnings distribution. The portion of the total household income inequality change accounted for by each factor is thus done in isolation with the same starting conditions. However, using this approach, the sum of the changes in household income inequality accounted for by all factors may overstate or understate the true change in inequality since that time.

While this paper uses the former approach, adding changes from factor components sequentially, the main results are consistent with those found using the approach of analyzing each factor in isolation. The choice between these methods alters the magnitude of some results but they both present a similar picture of the factors accounting for the rise in income inequality. The results in the following section focus exclusively on the primary sequential analysis, but the results using the alternate approach are provided in Appendix Table 2.1.

Robustness of results to changes in inequality index. While the Gini coefficient is the primary focus of this paper, a similar decomposition can be performed for other inequality measures. This including the three Generalized Entropy indices which, similar to the Gini, satisfy each of the desirable properties of inequality indices described previously. When performing the decomposition on these alternate inequality measures, the results are largely consistent with those found for the Gini coefficient. Results for the GE(0), the Mean Log Deviation; GE(1), the Theil index; and GE(2), half of the squared coefficient of variation, are provided in Appendix Tables 2.2 through 2.4.

2.5 Decomposition Results

In order to explain factors accounting for trends in household income inequality, it is necessary to first observe the trends being explained. The first row of Table 2.2 does just that – providing the average annual percentage change in Gini coefficients for each business cycle since 1967. From Row 1 of Table 2.2, it is apparent that inequality fell slightly in the late 1960s business cycle (1967-1975), before rising dramatically in the late 1970s business cycle (1975-1983). It continued to rise, although at a somewhat moderated pace, in the late 1980s business cycle (1983-1992). The year from 1992-

Table 2.2: Estimated average annual percentage change in the size-adjusted household income Gini coefficient attributable to factor components by business cycle

	1967-75	1975-83	1983-92	1992-93	1993-04	2004-07	1967-07
(1) Actual Gini Avg. Annual Pct. Change	-0.16	0.90	0.51	5.69	0.13	-0.23	0.42
Avg. Annual Pct. Change accounted for by:							
(2) Marriage Rates	0.20	0.20	0.12	0.20	0.05	0.08	0.13
(3) Male Employment Rates	0.22	0.09	0.00	-0.07	-0.02	-0.02	0.04
(4) Male Earnings Inequality	-0.05	0.32	0.31	4.10	0.09	-0.30	0.23
(5) Male Real Earnings Level	0.16	0.02	0.14	0.78	0.13	0.01	0.12
(6) Female Employment Rates	-0.04	-0.15	-0.19	-0.93	-0.08	0.00	-0.13
(7) Female Earnings Inequality	-0.20	0.05	0.08	0.81	0.02	0.12	0.03
(8) Female Real Earnings Level	0.06	0.01	0.04	-0.02	0.00	0.04	0.02
(9) Spouses' Earnings Correlation	0.03	0.24	0.10	0.01	-0.06	0.04	0.06
(10) Public Transfers Inequality	-0.23	0.06	0.02	0.06	0.03	0.01	-0.01
(11) Real Level of Public Transfers	-0.23	0.06	-0.08	-0.16	0.00	-0.02	-0.05
(12) Other Factors	-0.09	-0.01	-0.02	0.91	-0.03	-0.18	-0.02

Source: Authors calculations using Internal March CPS data (1967-2007)

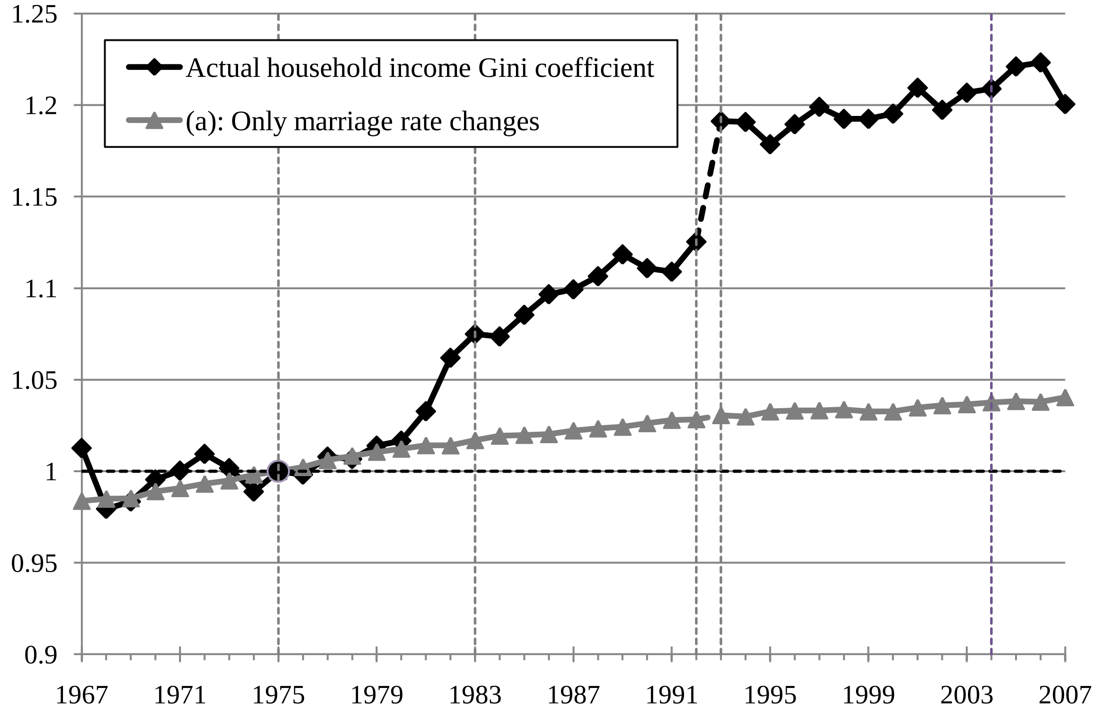
Note: 1992-1993 is separated from the 1992-2004 business cycle to separate the large artificial increase in inequality that occurred between 1992-1993 due to changes in the March CPS data collection procedures from actual changes occurring before and after that time.

1993 showed a substantial 5.69 percent increase in the Gini coefficient due to data collection changes in that year. A comparison of this increase to preceding and subsequent years highlights the impact of these changes and the importance of separating this year from other periods in the analysis.[26] In the remainder of the 1990s business cycle (1993-2004), the increase in inequality slowed even further compared to the periods of rapid growth in the early 1980s. Most recently, the beginning of the current business cycle (2004-2007) has then seen a fall in inequality – although the conclusions regarding the current business cycle are limited given that the complete business cycle cannot yet be observed.

The annual trends in income inequality can be observed in Figure 2.1 (with inequality in the base-year of 1975 normalized to 1). In this figure, it is apparent that income inequality trends are generally not steady over time. For example, the rapid rise in inequality from 1980-1983 was followed by a year of declining income inequality in 1984 before inequality continued to increase. The factors accounting for income inequality trends should explain not just the long-term trends but also these year-over-year fluctuations. Thus, it is valuable to consider both the long-term inequality changes, which will be done across trough years of the business cycles, along with these annual fluctuations in income inequality. Having observed the trends in inequality to be accounted for, it is now possible to explore the factors underlying these trends.

Changes to the marriage rates. The first factor considered is the change in the fraction of households headed by married couples. Marriage rates have changed

[26] Further evidence that this one-year increase is artificial can be obtained by comparing March CPS results to other datasets. Burkhauser, et al. (2009) do so by showing that trends in top income shares for pre-tax, pre-transfer tax-unit income in the March CPS closely match Piketty and Saez's (2003) results using IRS tax records in most years. One year in which this comparison across datasets does not provide similar results is 1992-1993, where the top 1% income share increases substantially in the March CPS data due to the new Census Bureau procedures but are relatively constant in the IRS tax records.



Source: Authors calculations using Internal March CPS data (1967-2007)
Notes: Dashed vertical lines represent trough-years of each business cycle. Each series is dashed from1992-1993 due to the March CPS redesign of collection procedures that limits data comparability between 1992 and 1993. As a result of the redesign, comparisons of levels should not be made across years that span this period.
Figure 2.1: Estimated increase in Gini coefficient for household income when only marriage rates change and the income distribution is held constant (1967-2007), 1975 normalized to 1

rapidly over the past 40 years, as can be seen in Table 2.3 which provides the fraction of people living in households headed by a married couple in the trough year of each business cycle. In 1967, 82 percent of people lived in a household headed by a married couple. This percentage declined to just 63 percent in 2007.

This change could, in theory, increase or decrease income inequality depending on where unmarried individuals fall in the distribution. If unmarried individuals are concentrated near the middle of the distribution, the decline in marriage rates would decrease inequality. On the other hand, if they are more commonly found near either tail of the distribution, the reduction in marriage rates

Table 2.3: Percent of individuals living in a household with married, single-male, and single-female householders by year

	Married	Single Male	Single Female
1967	82.5	4.1	13.4
1975	77.4	5.8	16.8
1983	72.6	7.9	19.5
1992	67.5	10.3	22.2
1993	66.9	10.3	22.8
2004	63.6	13.1	23.4
2007	62.6	13.6	23.9

Source: Authors calculations using Internal March CPS data (1967-2007)
Notes: See Note to Table 2.1

would increase inequality. In general, single individuals are more likely to be living in poverty than their married counterparts. As a result, one would expect the decline in marriage rates to increase income inequality as it expands the population in the lower tail.

Starting with the March CPS sample from 1975, by using the reweighting technique described above it is possible to determine how much of the change in household income inequality can be accounted for by this decline in marriage rates. In doing so, note that the estimated effect focuses exclusively on changes to how many people marry, rather than who marries since it assumes the income distribution of married and single individuals remain unchanged. If, for example, there is an increase in assortative mating with high-income individuals increasingly marrying other high-income individuals, then the impact of such changes are not included here. Instead these changes will be observed in the analysis of changing correlations between male and female earnings. Thus, the impact of changing marriage rates measures exclusively the impact of the change in marriage rates across the population.

Row 2 of Table 2.2 shows the average-annual percentage change in household income inequality that is accounted for by marriage pattern changes in each business cycle. This can be compared to the actual changes in inequality observed for each business cycle presented in Row 1 of Table 2.2. In the late 1960s business cycle, when household income inequality fell by an average rate of -0.16 percent-per-year, the decline in marriage rates holding incomes for married and single individuals constant accounted for inequality increases of 0.20 percent per year. This suggests that had marriage patterns been unchanged during this period, the decline in income inequality would have been even greater.

During each subsequent business cycle through 2007 there continued to be a fall in marriage rates. In the late 1970s business cycles (1975-1983), declines in marriage rates accounted for approximately 23 percent of the total rise in inequality, as marriage rate declines accounted for a rise in inequality of 0.20 percent point per year compared to the actual increase of 0.90 percent per year. An identical 23 percent of the total increase in inequality can be explained by declines in marriage rates in the late 1980s (1983-1992) (0.12 percent per year increase due to marriage rates out of the total increase of 0.51 percent per year). During the 1990s (1993-2004) the decline in marriage rates slowed, thus reducing the rise in income inequality accounted for by marriage changes. But since income inequality growth slowed even more the marriage rate changes accounted for 40 percent of the 0.13 average annual percentage change in income inequality over the business cycle.

Figure 2.1 provides a comparison of the annual income inequality changes accounted for by the marriage rate changes and compares them to the actual income inequality changes seen during that time. When considered on an annual basis, rather than by business cycle, it is clear that changes to income inequality due to marriage rates are extremely steady compared to the actual changes in income inequality. There are wide fluctuations in the year-over-year changes in household income inequality, with inequality increasing or decreasing by as much as 3 percent in some years. In contrast, the inequality changes attributable to marriage rate changes are never more than 0.4 percent in any given year. This supports the view that while declines in marriage rates are contributing to the increase in income inequality, they generally cannot explain rapid shifts in inequality trends since they are a slow-moving factor.

Changes to the male employment rate. The second factor that may explain changes in household income inequality is the change in male employment rates. Previous studies that have considered the impact of changing employment rates on income inequality have only considered whether the individual is employed or not-employed (see, for example, Daly and Valetta 2006). However, not considering whether individuals work full time or part time will potentially miss important changes to work intensity over time. As such, this paper considers three potential employment statuses: working full time, working part time, and not-employed (both unemployed and not in the labor force).[27] It follows the previous literature, however, in focusing on the employment status and earnings of just the household head and his or her spouse rather than the employment of all individuals in the household. Thus all subsequent references to male earnings refer to the size-adjusted labor earnings of the male household head and references to female earnings refer to the size-adjusted labor earnings of the female household head.

It is likely that employment decisions of the household head are made conditional on marital status, since married individuals with other available sources of income likely have higher reservation wages and thus will enter the labor market at lower rates than single individuals. Therefore, to avoid confusing changes in income inequality resulting from changes in employment rates versus those resulting from changing marriage patterns, this analysis is performed conditional on marital status.

[27] Full time work is considered working at least an average of 35 hours per week for 50 weeks or more during the year. Part time work is considered working at least 1 hour for one or more weeks during the year, but working less than 35 hours per week or less than 50 weeks during the year.

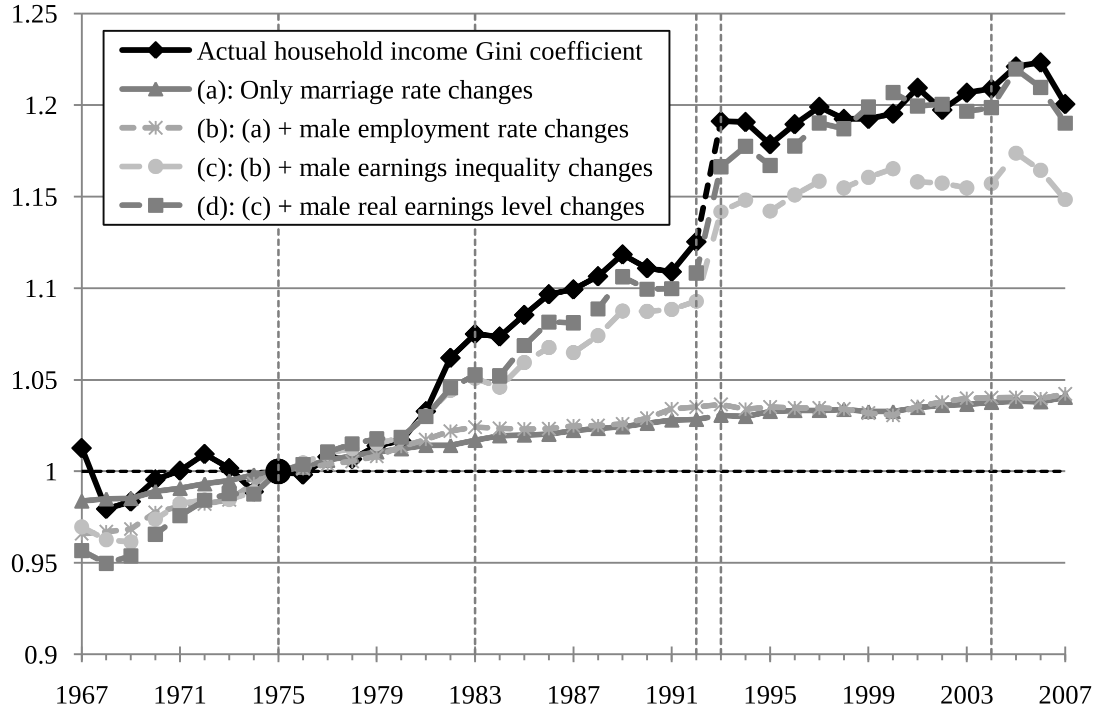
Table 2.4: Percent of individuals living in a household with a male householder working full-time, part-time, and not-working, given the marital status of the householder

	Percent living in a household with a married male householders working			Percent living in a household with a single male householders working		
	Full time	Part-time	None	Full time	Part-time	None
1967	76.2	17.0	6.9	52.0	24.8	23.2
1975	68.2	20.7	11.1	45.8	30.5	23.7
1983	64.0	21.4	14.6	50.6	28.4	21.0
1992	66.1	17.4	16.5	52.6	26.0	21.4
1993	66.9	16.3	16.8	54.3	25.0	20.8
2004	70.2	13.3	16.5	55.2	21.6	23.1
2007	70.8	12.3	16.9	56.8	21.2	22.0

Source: Authors calculations using Internal March CPS data (1967-2007)
Notes: See Note to Table 2.1

As was the case with the marital status of the head, employment rates of household heads have changed substantially over the past 30 years. The full-time and part-time employment rates for male household heads are provided in Table 2.4. Among married male household heads, the full-time employment rates fell substantially from 1967 through 1983 before recovering somewhat since then. Additionally, over the entire period there was an increase in non-working males as men shifted out of paid employment. Illustrating the importance of conditioning on marital status, the pattern for single males is quite different. The fraction of single-men not working in 2007 was very close to that seen in 1967. Among these men, most of the fluctuations in employment are in the form of shifts between full-time and part-time work.

The procedure for testing the impact of these changes in the employment rates is the same as that for evaluating the impact of changing marriage patterns, except that now a conditional rather than an unconditional reweighting procedure is used. Once again, the income of all individuals remains unchanged conditional on marital status



Source: Authors calculations using Internal March CPS data (1967-2007)
Notes: See Note for Figure 2.1

Figure 2.2: Estimated increase in the household income Gini coefficient resulting from male employment status and labor earnings changes (1967-2007), 1975 normalized to 1

and employment status.

In Row 3 of Table 2.2, the additional impact of male employment rate changes are presented for each business cycle. Although male employment changes substantially increased inequality in the late 1960s and were still important over the late 1970s business cycles, these changes have had very little impact on inequality since that time.

This is further evidenced by comparing Series (b) of Figure 2.2, which presents inequality trends when both marriage rates and male employment rates change, to Series (a) of Figure 2.2 when only marriage rates changed. With the exception of the period from 1967 to 1975, where the male employment changes clearly led to a further increase in inequality, the inequality trends in these two series are quite similar. Thus, male employment rate changes offer relatively little additional information regarding the changes in inequality seen since 1975. To the extent that changes in male earnings impacted household income inequality, it is evident that these changes have generally occurred among workers with a given employment status rather than due to shifts in intensity of employment.

Changes to the male labor earnings distribution.[28] Given that the changes in male employment rates generally had little impact household income inequality, to what extent did changes to the male earnings distribution impact household income inequality? Since male earnings inequality rose substantially in the past 40 years, one would expect that male earnings changes had a large effect on household income inequality trends. Using the rank-preserving income exchange procedure previously described for testing the impact of changes to the income distribution, this section explores the extent to which this is the case. This is first done holding the conditional mean labor earnings constant at 1975 levels, concentrating only on changes to male earnings inequality and excluding the impact of the real wage growth.

In Row 4 of Table 2.2, it can be seen that male earnings inequality explains only a small fraction of the decline in income inequality between 1967 and 1975. These changes in male earnings inequality explain a 0.05 average annual percent decrease in inequality over this business cycle, which is less than one-fourth of the 0.22 average annual percent increase in inequality seen from male employment changes over the same period. Thus, male earnings and employment changes alone cannot account for the small decline in household income inequality before 1975.

The importance of changes in male earnings inequality is very different after 1975, however. In the late 1970s business cycle male earnings inequality changes

[28] References to the male and female earnings distribution refer to the combination of changes to male labor earnings inequality and to changes in the real level of male earnings.

account for a 0.32 average annual percentage increase in inequality, 35 percent of the 0.90 actual average annual percentage increase. In both the late 1980s and 1990s business cycles it accounts for over 60 percent of the observed increases in inequality.

When the real male earnings growth is included along with the male earnings inequality changes, it is evident that complete shifts in the male earnings distribution are even more important for explaining household income inequality trends.[29] In the late 1980s and late 1990s business cycles, the inclusion of real male earnings growth explains an additional 0.14 and 0.13 percent per year of the inequality growth respectively (Row 5 of Table 2.2).

Additionally, when considering the annual trends in income inequality explained by including male earnings inequality (Series (c) of Figure 2.2) and male real earnings growth (Series (d) of Figure 2.2) in addition to the marriage rate and male employment changes, one can observe that many of the year-over-year fluctuations in inequality can largely be explained by the inclusion of these factors. While the magnitudes of inequality changes do not always match those actually observed, the years in which inequality increase and decrease attributable to male earnings, employment, and real income changes are generally consistent with the actual income inequality series.

The combination of male employment rate changes, male earnings inequality changes, and male real earnings growth are clearly extremely important for understanding the long-term changes in income inequality that have occurred over the past 40 years. However, it is also notable that the relationship between these male-earnings factors and household income inequality trends has changed over time. In the

[29] Real earnings are inflation adjusted using the CPI-U-RS series. This series is used by the Census Bureau for their historical income series (Denavas-Walt, Proctor, and Smith, 2009) and incorporates recent improvements in the CPI to provide a more consistent inflation series than that provided by the unadjusted CPI. For more details on the CPI-U-RS series, see Stewart and Reed (1999).

late 1970s business cycle, the sum of the contributions of these three factors explains a 0.43 average annual percentage increase in income inequality (Rows 3 through 5 of Table 2.2), or about one-half of the net increase in income inequality during this period. In the late 1980s business cycle, they combined to explain a 0.45 average annual percentage increase in income inequality, which is 88 percent of the actual income inequality growth observed in this business cycle.

By the 1990s business cycle, while the combined contribution of these factors slowed to explain a 0.19 average annual percentage increase, this slowdown is less substantial than the actual slowdown in income inequality that occurred. As a result, in the 1990s business cycle the male earnings and employment changes accounted for 143 percent of the actual increase in inequality. This suggests that had it not been for other factors mitigating the inequality growth from male employment and earnings changes, household income inequality growth would not have slowed to the extent that it did in the 1990s.

More broadly, since Figure 2.2 illustrates that male employment and earnings alone understate the inequality increases in the 1970s and 1980s and overstates the increase in the 1990s, had other changes beyond male employment and earnings changes not been occurring, household income inequality growth would have been slower in the late 1970s and 1980s business cycles but faster in the 1990s. Therefore, to reconcile the differences in the timing of the inequality increases, it is necessary to also consider the changes to the earnings distributions of other income sources and changes to the correlation between male and female earnings.

Changes to female employment rates and the female earnings distribution. It has been well documented that female employment rose dramatically over the past 40 years. Unlike the observation for male employment where married men exited the work force while single men increased their employment rates, females in 2007 are

Table 2.5: Percent of individuals living in a household with a female-householder working full-time, part-time, and not-working, given the marital status of the householder

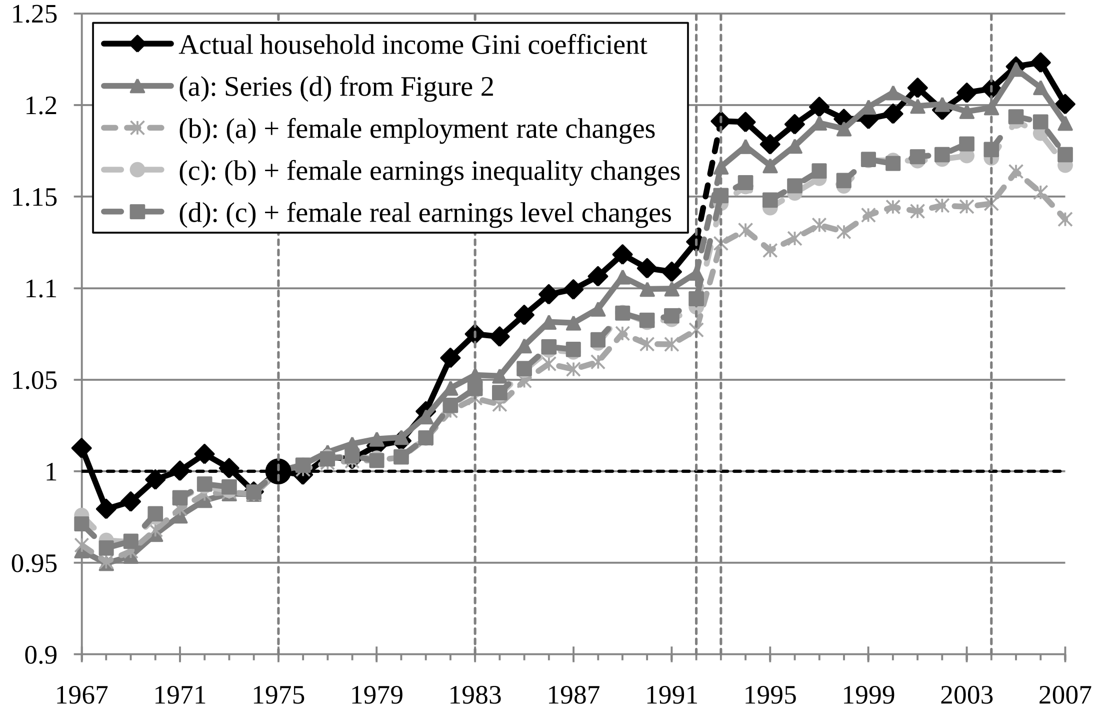
	Percent living in a household with a married female householders working			Percent living in a household with a single female householders working		
	Full time	Part-time	None	Full time	Part-time	None
1967	18.3	30.9	50.8	30.1	27.9	42.1
1975	20.2	32.0	47.8	28.6	26.0	45.4
1983	26.0	32.8	41.2	32.7	23.2	44.1
1992	35.4	30.5	34.2	36.9	22.5	40.6
1993	34.9	31.3	33.7	36.3	23.4	40.2
2004	38.9	26.2	35.0	42.7	23.1	34.2
2007	41.3	24.4	34.3	43.3	22.5	34.2

Source: Authors calculations using Internal March CPS data (1967-2007)
Notes: See Note to Table 2.1

both more likely to be employed and more likely to work full-time than they were in 1967 regardless of their marital status. This is illustrated in Table 2.5, which shows a 23 percentage point increase in full-time employment among married women and a 13 percentage point increase in full-time employment among single women.

Using the same reweighting procedure as above for evaluating the impact of changing male employment rates on household income inequality, Row 6 of Table 2.2 illustrates that the rise in female employment slowed the rise in income inequality over the past 40 years. During the late 1970s and 1980s business cycles this reduction in inequality growth was more substantial given the rapid growth in female employment during that period. If females entering the labor market had the same earnings profiles of women in 1975, increases in female employment would have reduced inequality growth by -0.15 and -0.19 percent per year in the 1970s and 1980s business cycles respectively. This is quite different from the trivial impact of the male employment rate changes on household income inequality over this period.

This can be seen more clearly in Figure 2.3. Series (a) of Figure 2.3 reproduces



Source: Authors calculations using Internal March CPS data (1967-2007)
Notes: See Note for Figure 2.1

Figure 2.3: Estimated increase in the household income Gini coefficient resulting from female employment status and labor earnings changes (1967-2007), 1975 normalized to 1

Series (d) of Figure 2.2 – the change in inequality that would have occurred if only marriage rates and the male employment rate and earnings distribution changed. Series (b) of Figure 2.3 then shows the change in inequality if marriage rates, the male employment rate and earnings distribution, and female employment rates changed. Considering the difference between Series (b) and Series (a), the marginal effect of female employment changes did considerably reduce the increases in inequality – as expected from the business cycle results described above. However, like changes in marriage rates, this change was mainly important for long-term trends and hence its inclusion did not have much effect on the years in which inequality increased and the years in which it decreased.

More importantly, just as the distribution of earnings among male workers was changing, the conditional female earnings distribution changed as well. Series (b) of Figure 2.3 does not incorporate these changes to the female earnings distribution. To examine how changes to the female earnings distribution of working women contributed to household income inequality changes, the conditional rank-preserving income exchange procedure is used once again.

As can be seen in Row 7 of Table 2.2, including the changes to female earnings inequality, but holding the conditional mean female incomes constant offsets the household income inequality declines since 1975 that resulted from the increase in female employment. However, the net increase in income inequality since 1975 is still smaller than it would have been had both female employment and female earnings inequality been unchanged at the 1975 levels. Prior to 1975, the contribution of changing female earnings inequality was different. During the late 1960s business cycle, the decline in female earnings inequality reinforced the slower inequality growth from women entering the labor market and these two factors account for a decline in income inequality during this period. However, the combination of factors considered thus far still suggest an increase in inequality from 1967-1975 rather than the slight decrease that actually occurred.

Unlike the case for men, Row 8 of Table 2.2 shows the increase in inequality from including the real female earnings growth in addition to the female earnings inequality changes. In each business cycle since 1975, the average annual percentage increase in inequality attributable to real earnings growth of female earnings was less than 0.05 percent per year. Thus the combined effect of female employment and earnings distribution changes (Rows 6 through 8 of Table 2.2) reduced inequality in each business cycle through the 1990s. In the 1970s and 1980s business cycles, these combined factors reduced inequality growth by approximately 10 percent of the actual increase, and in the 1990s it reduced inequality growth by approximately 40 percent of the now slower actual increase.

The year-over-year changes in inequality including these factors (Series (c) and Series (d) of Figure 2.3) present a similar picture to that described above when female employment rates changed but the conditional female earnings distributions were held constant. The increase in inequality is slower than that seen when both female employment and earnings distributions were held constant but the year-over-year patterns in inequality are only slightly effected.

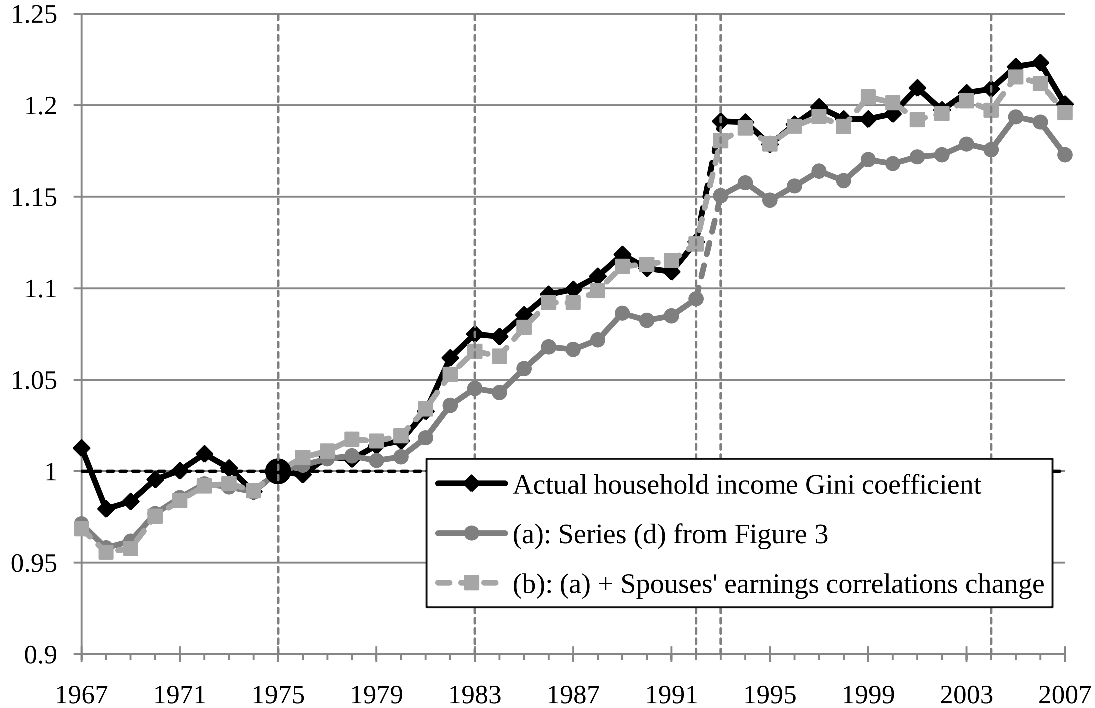
Changes to spouses' earnings correlations. Thus far the analysis has assumed that the rank correlation across income sources remains unchanged. However, this has not been the case. The correlation between male and female labor earnings has risen substantially over the past 30 years, which could occur for several reasons. These include an increase in assortative mating or from high-skill women who were married to high-earning men and thus chose not to work in the 1970s have entered the labor market at a disproportionately high rate and now have positive labor earnings. This increase in the correlation of earnings between male and female household heads has the potential to greatly increase household income inequality by concentrating wealth into a smaller number of households. Since correlation changes can result both from shifts in the correlation of spouses' employment decisions and from shifts in the correlation of earnings when both spouses are employed, the analysis of correlation changes only conditions on the household heads' marital status and not their employment status.

Previously, it was observed that almost all of the total increase in household income inequality could be explained by the combination of changes to the marital status of the household head and the employment status and earnings distributions of male household heads. However, as was seen in Figure 2.2 these factors alone led to an understatement of the household income inequality increase in the late 1970s and

1980s business cycles and an overstatement of the increase in the 1990s. This was particularly noticeable during the late 1970s business cycle where income inequality was rising quite rapidly but male employment and earnings distribution changes could explain less than half of this increase.

As is evident in Row 9 of Table 2.2, these timing differences can largely be explained by the change in earnings correlation. In the late 1970s and 1980s business cycles, the change in earnings correlations were positively associated with household income inequality changes. The increase in correlation accounted for a 0.24 percent-per-year increase in household income inequality in the late 1970s business cycle and a 0.10 percent-per-year increase in the 1980s business cycle. As a result, the rise in household income inequality over these two business cycles outpaced the rise attributable to only male employment and earnings distribution changes.

This reversed in the 1990s business cycle. Starting in the 1990s, the effect of changes in spouses‘ rank correlation of labor earnings reversed. Rather than accounting for an increase in household income inequality, as occurred in the 1970s and 1980s, changes in spouses’ earnings rank correlations mitigated the increase in household income inequality in the 1990s. Had earnings rank correlations been unchanged, the rise in household income inequality in the 1990s would have been 0.06 percent-per-year higher, a 39 percent increase over the observed inequality increase. Thus, as a result of changing earnings correlations and the reduction in household income inequality attributable to female employment rate changes, the actual rise in household income inequality in the 1990s was slower than that attributable to only male earnings distribution and employment rate changes. In the beginning of the current business cycle, from 2004-2007, however, correlation between male and female earnings began growing slightly again so reductions in their correlations should not necessarily be expected to continue holding down inequality increases in the



Source: Authors calculations using Internal March CPS data (1967-2007)
Notes: See Note for Figure 2.1
Figure 2.4: Estimated increase in the household income Gini coefficient resulting from spouses' earnings rank-correlation changes (1967-2007), 1975 normalized to 1

future.

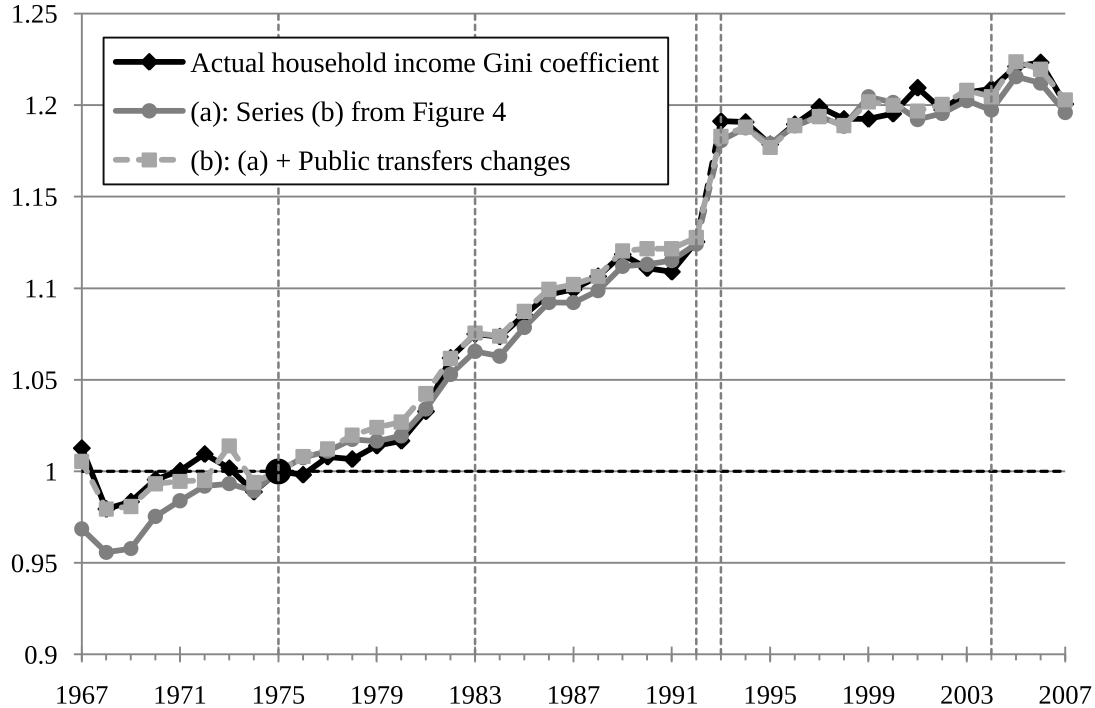
The impact of these correlation changes can also be seen in Series (b) of Figure 2.4. When spouses’ earnings correlations are changed along with marriage rates, male and female employment rates, and male and female earnings distributions, the explained inequality trend very closely matches the actual change in inequality since 1975. Thus, while there are other factors that could also influence household income inequality, such as changes in public transfers, non-labor income, or labor earnings for non-household heads, these other factors play only a small role in explaining inequality changes from 1975-2007.

However, it is also apparent in Figure 2.4 that the previously discussed factors are less successful at explaining inequality changes prior to 1975. Changing only these factors suggests that household income inequality would have increased by 3.1 percent over this 8 year period, when in fact, household income inequality declined by 1.3 percent. Therefore, there must be another factor that accounts for why household income inequality was not increasing. This factor is public transfers. Between 1967 and 1975 there were sizeable increases in public transfers for those near the bottom of the income distribution. As a result, excluding these changes to public assistance programs ignores a source of growing income for individuals at the bottom of the income distribution during the late 1960s business cycle.

It is possible that these increases in public transfers also induced individuals to reduce their levels of employment and contributed somewhat to the rise in inequality in the late 1960s observed when only changes to employment, labor earnings, and marriage rates were considered. However, even if that is the case, focusing only on labor earnings changes and ignoring changes in public transfers income will lead to a misstatement of shifts in the well-being of individuals at lower-tail of the distribution and of changes in household income inequality.

Public Transfers. The impact of changes to the distribution of public transfers can be added by using a rank-preserving income exchange as was done for both male and female earnings. The changes in the inequality of public transfers distributions that occurred in the 1960s business cycle accounted for a -0.23 average annual percentage change in the Gini coefficient during this period (Row 10 of Table 2.2). Furthermore, the increase in mean public transfers over this period accounted for an additional -0.23 average annual percentage change in the Gini coefficient (Row 11 of Table 2.2). As a result, public transfers changes explain why the earlier analysis expected an increase in income inequality in the late 1960s business cycle while inequality actually decreased slightly.

The influence of including public transfers can be seen more clearly by



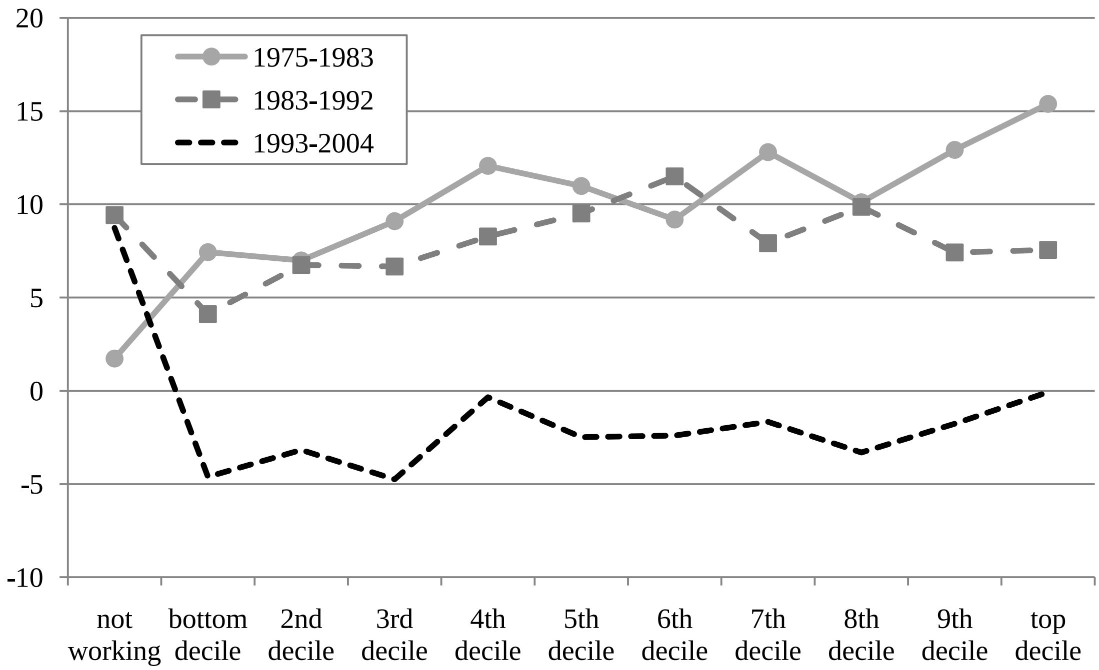
Source: Authors calculations using Internal March CPS data (1967-2007)
Notes: See Note for Figure 2.1

Figure 2.5: Estimated increase in the household income Gini Coefficient resulting from public transfers changes (1967-2007), 1975 normalized to 1.

comparing Series (b) of Figure 2.5, which includes these public transfers changes along with all previously discussed factors, to Series (a) of Figure 2.5 which excludes the public transfers changes. While changes to public transfers had very little additional impact on inequality after 1975, their inclusion largely reconciles the unexplained changes in income inequality prior to 1975.

2.7 Understanding the trend in the correlation of spouses' earnings

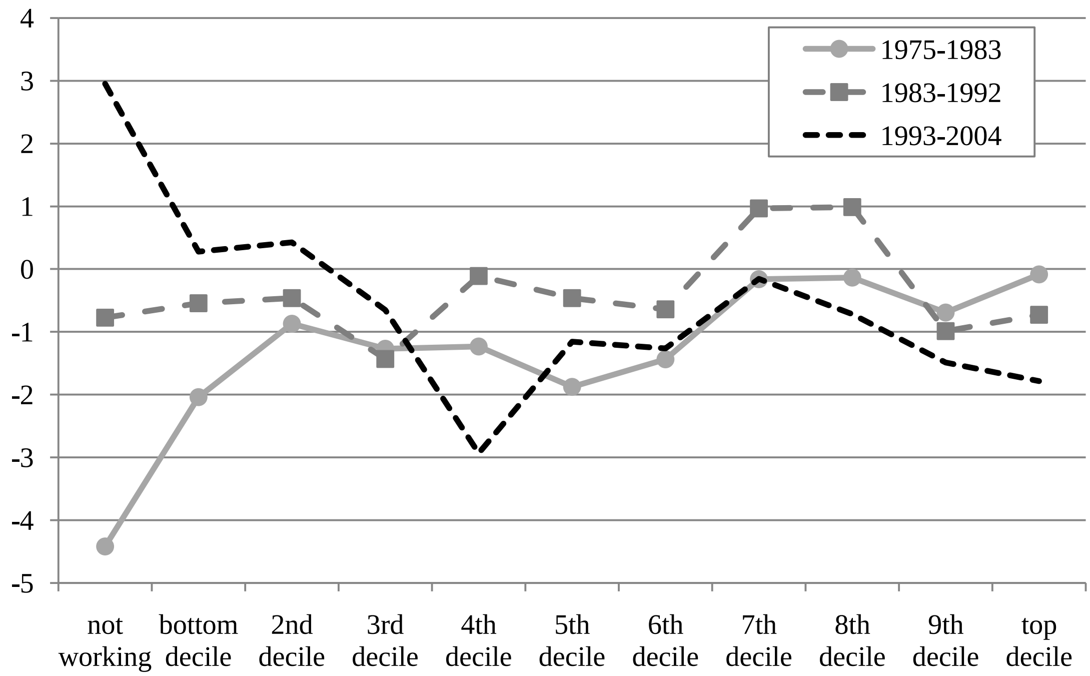
Given the importance that the correlation between spouses' earnings had on the trends in household income inequality, it is worth considering what led to the rise in correlation in the 1970s and 1980s and the decline in correlation in the 1990s. The changes in correlation can come either from shifts in the correlation of earnings among



Source: Authors calculations using Public Use March CPS data (1967-2007)
Figure 2.6: Percentage change in female employment among married, working age women by decile of their husband's labor earnings

couples where both individuals work or from shifts in the frequency of two-earner couples at different points in the income distribution. While not eliminating the possibility of the former explanation, there is evidence that changing employment patters are a key element of the correlation trends.

Figure 2.6 illustrates the change in female employment rates among married working age women in each full business cycle since 1975 based on the decile of their husband's labor earnings. When considering employment trends in this way, it is apparent that in the 1975-1983 business cycle, the most rapid rise in female employment occurred among women married to high earning men. Among working-age women married to men in the top decile of the male earnings distribution, employment increased by over 15 percentage points. However, among working-age women married to non-working men employment only increased by 2 percentage



Source: Authors calculations using Public Use March CPS data (1967-2007)
Figure 2.7: Percentage change in male employment among married, working age women by decile of their wife's labor earnings

points. This difference is important in understanding the increase in correlations since the increases in female employment among high earning men must increase the spouses' earnings correlation as two uncorrolated observations (a high earning man married to a woman with no earnings) now become more correlated (a high earning man married to a woman with some positive earnings).

Looking at the 1993-2004 business cycle when correlations declined, the pattern is quite different. During this period, women married to non-working men increased their employment by almost 10 percent while women married to working men in any decile saw a decline in employment rates. This would therefore lead to a decline in correlations as correlated earnings (two non-earners) become less correlated with the woman entering the labor market.

Figure 2.7 provides a similar analysis of the change in male employment rates among married working age men based on the decile of their wife's labor earnings. This analysis finds similar results to that of female employment trends. In the 1975-1983 business cycle the slowest decline in male employment is among men who are married to high-earning women. But in the 1993-2004 business cycle the trend is reversed with only men married to non-working or low-earning women entering the labor market at a positive rate.

To gain even more insight into why there is a decline in couples with two non-workers in the 1990s requires further dividing the population into individuals in households with exactly two people and those in households with more than two people – which generally indicates children in the household. When doing so in Table 2.6, it is apparent that the increase in employment among working age women married to non-working men in the 1990s came entirely among women in households with three or more individuals. The same occurred among working age men married to non-working women, where the gains in employment were entirely among men in households with three or more individuals.

Although this analysis cannot provide information on precisely why the 1990s saw a decline in non-earning couples with children, one possible explanation is that it was in response to changes in public policies during that time. Significant expansions of the Earned Income Tax Credit that occurred between 1993 and 1996 provided strong incentives for low income households with children to have at least one individual working in the labor market. Additionally, declines in the welfare program made it more difficult for households with children to survive on government benefits alone. Thus, employment patterns may reflect responses to public policies and the influence of these policies on income inequality trends go beyond that seen for labor earnings through their impact on the correlation of spouses' earnings.

Table 2.6: Correlation coefficient between size-adjusted male and female household head labor earnings

	Percent living in a household with a married female householders working			Percent living in a household with a single female householders working		
	Full time	Part-time	None	Full time	Part-time	None
1967	18.3	30.9	50.8	30.1	27.9	42.1
1975	20.2	32.0	47.8	28.6	26.0	45.4
1983	26.0	32.8	41.2	32.7	23.2	44.1
1992	35.4	30.5	34.2	36.9	22.5	40.6
1993	34.9	31.3	33.7	36.3	23.4	40.2
2004	38.9	26.2	35.0	42.7	23.1	34.2
2007	41.3	24.4	34.3	43.3	22.5	34.2

Source: Authors calculations using Internal March CPS data (1967-2007)
Notes: See Note to Table 2.1

2.6 Conclusions

Numerous studies have documented the rapid rise in male earnings inequality and household income inequality that occurred during the 1980s, and the slower growth in inequality for each of these series in the 1990s. Despite the relative similarities in the overall increases in inequality using these two income definitions, the relationship between these series is not one-to-one and there are numerous other factors that should also be considered to understand trends in household income inequality.

When disaggregating the increase in household income inequality into its component sources, results for the 1970s and 1980s are consistent with those found in earlier studies by Burtless (1999) and Burtless and Karoly (1995). Male earnings and employment changes account for just 47 percent of the rise in household income inequality in the late 1970s business cycle – which also includes the early years of the 1980s. While male earnings inequality growth has slowed since that time, it has not slowed as markedly as household income inequality growth. As a result, the importance of male employment and earnings distribution changes have grown in importance for explaining household income inequality trends and accounted for 143 percent of the increase in household income inequality in the 1990s. This indicates that had it not been for other factors mitigating the increase in inequality from male earnings then household income inequality growth would not have slowed as substantially in the 1990s. The additional reduction in household income inequality growth in the 1990s was in large part due to changes in spouses' earnings correlations which reduced household income inequality growth during the decade by 39 percent. This is in marked contrast to the earlier business cycles when increasing correlations between male and female earnings accelerated the growth in income inequality. However, it also appears to be coming from a different area of the income distribution – with the earlier rise in correlations coming from an increase in two-earner couples while the decline in the 1990s appears to be coming from a decline in no-earner couples.

An additional factor that changed dramatically over the past 40 years that influences income inequality is the rate of female employment. Consistent with the findings by Daly and Valetta (2006), the increase in female employment led to a reduction in household income inequality. However, this reduction was mitigated by increases in female earnings inequality among working women. Nevertheless, the combined impact of female employment rate changes and female earnings distribution changes did slow the increase in household income inequality by 10 percent in the 1970s and 1980s business cycles and by 40 percent in the 1990s.

Income sources other than labor earnings also have influenced income inequality at times during the past 40 years. During the business cycle spanning the late 1960s and early 1970s, increases in public transfer income at the bottom of the income distribution were offsetting the rises in inequality accounted for by other factors. While changes in public transfers have had little influence on trends in income inequality since that time, their large effect during this early business cycle highlights the importance of considering income beyond just labor earnings when evaluating levels and trends of income inequality.

The remaining contributor to changing household income explored was the decline in marriage rates over the past 40 years. Other than changes to male earnings inequality, this decline was the most important factor contributing to the rise in household income inequality since 1975 and accounts for a steady increase in income inequality since that time. Thus, it is evident that demographic shifts play an important role alongside changes in labor-market compensation in explaining the long-term trends in income inequality in the United States.

APPENDIX

Appendix Table 2.1: Estimated Gini increase attributable to factor components using alternate method of analysis rather than sequential analysis

	1967-75	1975-83	1983-92	1992-93	1993-04	2004-07	1967-07
(1) Actual Gini Avg. Annual Pct. Change	-0.16	0.90	0.51	5.69	0.13	-0.23	0.42
Avg. Annual Pct. Change accounted for by:							
(2) Marriage Rates	0.20	0.20	0.12	0.20	0.05	0.08	0.13
(3) Male Employment Rates	0.25	0.13	0.02	-0.04	-0.02	-0.01	0.07
(4) Male Earnings Inequality	-0.04	0.36	0.37	4.94	0.11	-0.24	0.28
(5) Male Real Earnings Level	0.21	0.03	0.13	0.49	0.05	-0.03	0.10
(6) Female Employment Rates	-0.06	-0.20	-0.17	-0.03	-0.04	-0.08	-0.10
(7) Female Earnings Inequality	-0.20	0.05	0.05	0.65	0.00	0.10	0.01
(8) Female Real Earnings Level	0.01	0.01	0.07	0.19	0.08	0.06	0.05
(9) Spouses’ Earnings Correlation	0.02	0.24	0.11	0.06	-0.05	0.06	0.06
(10) Public Transfers Inequality	-0.23	0.05	0.01	0.06	0.05	0.02	-0.01
(11) Real Level of Public Transfers	-0.38	0.04	-0.10	-0.19	-0.01	-0.04	-0.10
(12) Other Factors	0.07	-0.01	-0.10	-0.64	-0.09	-0.13	-0.07

[1]Since other factors are the residual change, they are not separated from the overestimation of inequality that results from applying the shift-share analysis on 1975 data for all factors rather than applying the analysis sequentially through each factor.
Source: Authors calculations using Internal March CPS data (1967-2007)
Notes: See Note to Table 2.1

Appendix Table 2.2: Estimated average annual percentage change in the size-adjusted household income GE(0) coefficient attributable to factor components by business cycle

	1967-75	1975-83	1983-92	1992-93	1993-04	2004-07	1967-07
(1) Actual Gini Avg. Annual Pct. Change	-0.33	2.07	0.48	11.79	1.06	-1.03	0.96
Avg. Annual Pct. Change accounted for by:							
(2) Marriage Rates	0.48	0.46	0.25	0.41	0.10	0.14	0.27
(3) Male Employment Rates	0.35	0.12	0.00	-0.10	-0.02	-0.04	0.06
(4) Male Earnings Inequality	0.06	0.71	0.11	7.15	0.03	-1.36	0.23
(5) Male Real Earnings Level	0.36	0.04	0.28	1.62	0.24	-0.01	0.24
(6) Female Employment Rates	0.00	-0.26	-0.30	-1.70	-0.20	0.08	-0.21
(7) Female Earnings Inequality	-0.38	0.03	0.06	1.10	-0.01	0.30	0.01
(8) Female Real Earnings Level	0.09	0.01	0.10	0.07	0.01	0.10	0.05
(9) Spouses' Earnings Correlation	-0.07	1.08	0.21	1.21	-0.12	-0.43	0.17
(10) Public Transfers Inequality	-0.91	0.81	0.14	0.52	1.34	0.04	0.47
(11) Real Level of Public Transfers	-0.56	0.15	-0.17	-0.28	0.01	-0.06	-0.11
(12) Other Factors	0.26	-1.10	-0.20	1.80	-0.33	0.20	-0.28

Source: Authors calculations using Internal March CPS data (1967-2007)

Notes: See Note to Table 2.1

Appendix Table 2.3: Estimated average annual percentage change in the size-adjusted household income GE(1) coefficient attributable to factor components by business cycle

	1967-75	1975-83	1983-92	1992-93	1993-04	2004-07	1967-07
(1) Actual Gini Avg. Annual Pct. Change	-0.69	1.53	1.28	19.89	0.26	-1.26	0.92
Avg. Annual Pct. Change accounted for by:							
(2) Marriage Rates	0.37	0.38	0.20	0.30	0.08	0.12	0.21
(3) Male Employment Rates	0.38	0.16	0.00	-0.11	-0.03	-0.03	0.07
(4) Male Earnings Inequality	-0.23	0.39	0.92	16.86	0.16	-1.18	0.62
(5) Male Real Earnings Level	0.33	0.04	0.28	2.70	0.28	-0.06	0.28
(6) Female Employment Rates	-0.11	-0.34	-0.47	-4.23	-0.14	0.16	-0.32
(7) Female Earnings Inequality	-0.41	0.11	0.18	2.92	0.02	0.22	0.10
(8) Female Real Earnings Level	0.06	0.00	0.03	-0.31	-0.05	0.17	0.00
(9) Spouses' Earnings Correlation	0.03	0.56	0.25	0.57	-0.11	0.20	0.15
(10) Public Transfers Inequality	-0.50	0.17	0.05	-0.17	0.12	0.01	-0.01
(11) Real Level of Public Transfers	-0.44	0.13	-0.16	-0.32	0.01	-0.03	-0.09
(12) Other Factors	-0.17	-0.06	-0.01	1.69	-0.08	-0.84	-0.12

Source: Authors calculations using Internal March CPS data (1967-2007)

Notes: See Note to Table 2.1

Appendix Table 2.4: Estimated average annual percentage change in the household income GE(2) coefficient attributable to factor components by business cycle

	1967-75	1975-83	1983-92	1992-93	1993-04	2004-07	1967-07
(1) Actual Gini Avg. Annual Pct. Change	-1.70	0.77	2.21	52.84	0.03	-4.06	1.34
Avg. Annual Pct. Change accounted for by:							
(2) Marriage Rates	0.33	0.38	0.21	0.23	0.06	0.09	0.17
(3) Male Employment Rates	0.38	0.19	0.00	-0.09	-0.03	-0.03	0.06
(4) Male Earnings Inequality	-0.80	-0.19	2.05	50.71	0.05	-3.64	1.30
(5) Male Real Earnings Level	0.36	0.03	0.32	8.04	0.31	-0.54	0.40
(6) Female Employment Rates	-0.24	-0.51	-0.82	-12.92	-0.09	0.98	-0.54
(7) Female Earnings Inequality	-0.53	0.15	0.30	7.47	-0.03	0.04	0.20
(8) Female Real Earnings Level	-0.04	-0.05	-0.14	-1.83	-0.23	0.51	-0.12
(9) Spouses' Earnings Correlation	0.01	0.71	0.53	5.15	-0.22	0.71	0.32
(10) Public Transfers Inequality	-0.54	0.07	0.03	-1.44	0.11	0.01	-0.05
(11) Real Level of Public Transfers	-0.36	0.13	-0.15	-0.49	0.00	-0.01	-0.07
(12) Other Factors	-0.27	-0.14	-0.11	-1.98	0.07	-2.18	-0.44

Source: Authors calculations using Internal March CPS data (1967-2007)
Notes: See Note to Table 2.1

REFERENCES

Atkinson, Anthony B. 1998. *Poverty in Europe.* Oxford, UK: Blackwell.

Atkinson, Anthony B. and Andrea Brandolini. 2001. “Promises and Pitfalls in the Use of Secondary Data Sets: Income Inequality in OECD Countries as a Case Study.” *Journal of Economic Literature*, 39(3): 771–799.

Atkinson, Anthony B., Lee Rainwater, and Timothy Smeeding. 1995. *Income Distribution in OECD Countries. Evidence from the Luxembourg Income Study*. Social Policy Studies No. 18. Paris: Organization for Economic Cooperation and Development.

Autor, David H., Lawrence F. Katz, and Melissa S. Kearney. 2008. “Trends in U.S. Wage Inequality: Revising the Revisionists”. *Review of Economics and Statistics* 90(2): 300–323

Bayez, Gulgun and Kenneth A. Couch. 2008. “Demographic Contributions to Income Inequality of the Working age Population: A Decomposition Analysis by Income Source from 1969-2006.” Working Paper.

Berman, Eli, John Bound, and Zvi Griliches. 1994. “Changes in Demand for Skilled Labor within U.S. Manufacturing Industries.” *Quarterly Journal of Economics* 109(2): 367-398.

Buhmann, Brigitte, Lee Rainwater, Guenther Schmaus, and Timothy M. Smeeding. 1988. “Equivalence scales, well-being, inequality, and poverty: sensitivity estimates across ten countries using the Luxembourg Income Study (LIS) database.” *Review of Income and Wealth* 34(2): 115-142.

Burkhauser, Richard V., Shuaizhang Feng, and Stephen Jenkins. 2009. “Using a P90/P10 Ratio to Measure Inequality Trends with the Public Use Current Population Survey: A View from Inside the Census Bureau Vaults.” *Review of Income and Wealth* 55(1): 166-185.

Burkhauser, Richard V., Shuaizhang Feng, Stephen Jenkins, and Jeff Larrimore. 2009. "Recent Trends in Top Income Shares in the USA: Reconciling Estimates from March CPS and IRS Tax Return Data." NBER Working Paper No. w15320.

Burkhauser, Richard V., Shuaizhang Feng, Stephen P. Jenkins, and Jeff Larrimore. Forthcoming. "Trends in United States Income Inequality Using the March Current Population Survey: The Importance of Controlling for Censoring." *Journal of Economic Inequality.*

Burtless, Gary. 1999. "Effects of Growing Wage Disparities and Changing Family Composition on the US Income Distribution." *European Economic Review* 43(4-6): 853-865.

Burtless, Gary. 2009. "Demographic Transformation and Economic Inequality," in *The Oxford Handbook of Economic Inequality,* edited by Wiemer Salverda, Brian Nolan, and Timothy M. Smeeding. New York: Oxford University Press.

Cancian, Maria and Deborah Reed. 1998. "Assessing the Effects of Wives' Earnings on Family Income Inequality." *Review of Economics and Statistics* 80(1): 73-79.

Card, David and John E. DiNardo. 2002. "Skill-Biased Technological Change and Rising Wage Inequality: Some Problems and Puzzles." *Journal of Labor Economics* 20(4): 733-783.

Daly, Mary C. and Robert G. Valletta. 2006. "Inequality and Poverty in the United States: The Effects of Rising Dispersion of Men's Earnings and Changing Family Behavior." *Economica* 73(289): 75-98.

Del Boca, Daniela and Silvia Pasqua. 2003. "Employment Patterns of Husbands and Wives And Family Income Distribution in Italy (1977-98)." *Review of Income and Wealth* 49(2): 221-245.

Denavas-Walt, Carmen, Bernadette D. Proctor, and Jessica C. Smith, U.S. Census

Bureau. 2009. *Income, Poverty, and Health Insurance Coverage in the United States: 2008*. (August), Current Population Reports P60-235. Washington DC: GPO.

DiNardo, John E., Nicole Fortin, and Thomas Lemieux. 1996. "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semi-parametric Approach." *Econometrica* 64(5): 1001-1044.

Fei, John C. H., Gustav Ranis and Shirley W. Y. Kuo. 1978. "Growth and the Family Distribution of Income by Factor Components", *Quarterly Journal of Economics* 92(1): 17-53.

Feng, Shuaizhang, Richard V. Burkhauser, and J.S. Butler. 2006. "Levels and Long Term Trends in Earnings Inequality: Overcoming Current Population Survey Censoring Problems Using the GB2 Distribution." *Journal of Business and Economic Statistics* 24(1): 57-62.

Fournier, Martin. 2001. "Inequality Decomposition by Factor Component: A `Rank-Correlation' Approach Illustrated on the Taiwanese Case." *Recherches Economiques de Louvain – Louvian Economic Review* 67(4): 381-403.

Gottschalk, Peter and Sheldon Danziger. 2005. "Inequality of Wage Rates, Earnings and Family Income in the United States, 1975–2002." *Review of Income and Wealth* 51(2): 231–254.

Gottschalk, Peter and Timothy M. Smeeding. 1997. "Cross-National Comparisons of Earnings and Income Inequality." *Journal of Economic Literature,* 35(2): 633–687.

Jenkins, Stephen. 1995. "Accounting for Inequality Trends: Decomposition Analyses for the UK, 1971-86." *Economica* 62(245): 29-63.

Jenkins, Stephen. 1996. "Recent Trends in the UK Income Distribution: What Happened and why?" *Oxford Review of Economic Policy* 12(1): 29-46.

Jenkins, Stephen. and Van Kerm, Philippe 2009. “The Measurement of Income Inequality,” in Wiemer Salverda, Brian Nolan, and Timothy M. Smeeding (eds), *The Oxford Handbook of Economic Inequality*. New York: Oxford University Press.

Jones, Arthur F. and Weinberg, Daniel H. 2000. “The Changing Shape of the Nation’s Income Distribution.” *Current Population Reports,* U.S. Census Bureau, June 2000. Available online via: http://www.census.gov/prod/2000pubs/p60-204.pdf

Juhn, Chinhui, Kevin M. Murphy, and Brooks Pierce. 1993. “Wage Inequality and the Rise in Returns to Skill” *Journal of Political Economy* 101(3): 410-442

Karoly, Lynn A. and Gary Burtless. 1995. “Demographic Change, Rising Earnings Inequality, and the Distribution of Personal Well Being, 1959-1989.” *Demography* 32(3): 379-405.

Larrimore, Jeff, Richard V. Burkhauser, Shuaizhang Feng, and Laura Zayatz. 2008. “Consistent Cell Means for Topcoded Incomes in the Public Use March CPS (1975-2007).” *Journal of Economic and Social Measurement* 33(2-3): 89-128.

Lee, David S. 1999. “Wage Inequality in the United States During the 1980s: Rising Dispersion Or Falling Minimum Wage?” *Quarterly Journal of Economics* 114(3): 977-1023.

Lemieux, Thomas. 2006a. “Postsecondary Education and Increasing Wage Inequality.” *American Economic Review* 96(2): 195-199.

Lemieux, Thomas. 2006b. “Increasing Residual Wage Inequality: Composition Effects, Noisy Data, or Rising Demand for Skill?” *American Economic Review* 96(3): 461–498.

Lerman, Robert I. and Shlomo Yitzhaki. 1985. “Income Inequality Effects by Income Source: A New Approach and Applications to the United States.” *Review of*

Economics and Statistics 67(1): 151-156.

Mookherjee, Dilip, and Anthony Shorrocks. 1982. "A Decomposition Analysis of the Trend in UK Income Inequality." *Economic Journal* 92: 886-902.

Pasqua, Silvia. 2008. "Wives' Work and Income Distribution in European Countries." *European Journal of Comparative Economics* 5(2): 197-226.

Piketty, Thomas, and Emmanuel Saez. 2003. "Income Inequality in the United States, 1913–1998." *Quarterly Journal of Economics* 118(1): 1–39.

Ruggles, Patricia. 1990. *Drawing the Line: Alternative Poverty Measures and Their Implication for Public Policy*. Washington, DC: Urban Institute Press.

Ryscavage, Paul. 1995. "A surge in growing income inequality?" *Monthly Labor Review,* 118(8): 51–61.

Semega, Jessica and Ed Welniak. 2007. "Evaluating the impact of unrestricted income values on income distribution measures using the Current Population Survey's Annual Social and Economic Supplement (ASEC)". U.S. Census Bureau Research Report. Available online via: http://www.census.gov/hhes/www/income/Unrestricted-Income-Reporting.doc

Stewart, Kenneth J. and Stephen B. Reed. 1999. "CPI research series using current methods, 1978-98." *Monthly Labor Review* 122(6): 29-38.

Stigler, George J. 1946. "The Economics of Minimum Wage Legislation." *American Economic Review* 36(3): 358-365.

U.S. Census Bureau. 2008. *Current Population Survey (CPS) – Definitions and Explanations*. Available online via: http://www.census.gov/population/www/cps/cpsdef.html

CHAPTER 3

ARE THERE POSITIVE HEALTH EFFECTS FROM HIGHER INCOME? USING THE EARNED INCOME TAX CREDIT TO EXPLORE THE INCOME-HEALTH GRADIENT

Abstract

Although the existence of a positive relationship between income and morbidity has been well-documented in the literature, it is unclear whether the positive relationship exists because increased income allows individuals to purchase more health inputs contributing to their better health, because healthy individuals are more productive and can obtain higher wages in the labor market, or because a third factor contributes to increases in both health and income. This paper uses exogenous variation in income resulting from 17 years of changes in the generosity of state and federal Earned Income Tax Credit benefits to consider whether increases in income result in health improvements among the low income population. It finds only limited support for the theory that the relationship between income and morbidity is derived from shifts in income. In Survey of Income and Program Participation data, there is no evidence that increases in income improve self-reported health statuses and using March Current Population Survey data such evidence is only found when employment controls are excluded from the model. Additionally, while increases in income appear to reduce the prevalence of hearing and vision limitations when using corrective measures, it has no significant effect on other functional limitations considered.

3.1 Introduction

There is extensive evidence that a positive correlation exists between individuals' socioeconomic statuses and their health outcomes, with high income individuals tending to be in better health than those with low incomes. This relationship has been observed for a wide range of health measures including mortality (Backlund, Sorlie and Johnson, 1996; McDonough et al., 1997; Deaton and Paxson, 1998), chronic conditions (Case, Lubotsky and Paxson, 2002), obesity (Schmeiser, 2009), functional limitations (Zimmer and House, 2003), and self-reported health status (Deaton and Paxson, 1998). However, while it is generally accepted that this positive correlation exists, there is no clear consensus on the direction or the pathway of the relationship. For example, Case, Lubotsky and Paxson (2002) and Lindahl (2005) suggest that increased income improves health outcomes. Bound (1989), Haveman et al. (1995), and Smith (1999, 2004), on the other hand, suggest that lower incomes are due to the productivity declines that result from poor health and disabilities rather than the reverse. Still others, including Fuchs (1982), suggest that an outside factor, such as a high discount rate, could lead to both poor health and low incomes and create the observed correlation.

Distinguishing which pathway or pathways drive this relationship is extremely valuable for understanding the costs and benefits of public health policies and of public transfer programs. It has been suggested, for example, that increasing income transfer programs may be warranted if increases in income is demonstrated to have positive health effects (Deaton 2002, Lindahl 2005, and Herd, Schoeni and House 2008). But because there are multiple pathways which could lead to a correlation between health and income, it has proved difficult for researchers to determine the relationship’s causal direction.

One approach for examining the impact of income on health has been to focus on how family income impacts child health (Case, Lubotsky and Paxson, 2002). This relies on the assumption that children are not part of the labor force so their health does not impact family earnings. However, child health could still affect family income if the parents must take time and energy that could be devoted to labor market activities caring for the sick child. It is also possible that the impacts of income on health differ across the age distribution making it important to consider the impacts of income on adult health.

To further explore the direction of the income-health relationship requires using an exogenous variation in either health or income and evaluating the resulting impact of this variation on the other. Smith (1999, 2004) accomplishes this by using exogenous variation in health from the unanticipated onset of a chronic condition. He determines that health status has a causal impact on income and wealth among working age adults. However, the existence of a causal link in one direction does not preclude the existence of a link in the other direction as well.

The existing research using exogenous income variations to exploring the opposite direction of how individuals' incomes impact their health has had mixed findings. Comparing the health of mid-size lottery winners in Sweden to non-winners, Lindahl (2005) finds evidence that an individual's health status improves as a result of a positive income shock. However, in a comparison of mortality rates for individuals who higher Social Security benefits due to the Social Security notch, Snyder and Evans (2006) observed that individuals with higher benefits actually had higher mortality rates. Additionally, Schmeiser (2009) uses variation in income from EITC benefit changes to find that higher incomes appear to increase obesity rates among women. Supporters of the hypothesis that income influences health have criticized Snyder and Evans' findings, however, stating that much of the expected relationship occurs near the lower tail of the income distribution but that the Social Security Notch had a minimal impact on these low income individuals (Herd, Schoeni and House, 2008). While not directly measuring the effect of income on health, in related work

Ruhm (2000) considers how health statuses vary during business cycles and observes an improvement in health during economic downturns. Ruhm also notes that the theoretical expectations for the impact of increased income on health are not necessarily positive. This is because not all normal goods are positive health inputs. Thus, the higher income may also be used to purchase goods that are detrimental to health, such as alcohol or tobacco.

This paper adds to the literature exploring the effects of a variation in income on individuals' morbidity status. It does so by considering how an exogenous increase in income from a change in state-level Earned Income Tax Credit (EITC) benefits influences the health of individuals who are eligible for these benefits. Over the course of the 1990s, state and federal guidelines for EITC benefits changed dramatically and these changes in benefits led to substantial shifts in the incomes of low income families. These variations provide a natural experiment for exploring how changes in income influence the health of low income individuals.

There are three main benefits of using variation in EITC benefits for addressing the direction of the income-health gradient. First, previous research has suggested that the income-health gradient is strongest among the low income population (Backlund, Sorlie, and Johnson 1996, McDonough et al. 1997, Norris et al. 2003, Herd, Schoeni and House 2008). Since the EITC is targeted at the low income population, it can be used to analyze this group of particular relevance to the gradient. Second, because the EITC eligible population consists of working age adults, the impact of an exogenous shift in income can be observed for a younger population than previous studies that looked at changing benefits for retirees. Third, by examining morbidity among working age individuals, this paper can benefit from the large sample sizes of the March Current Population Survey (March CPS) data and the Survey of Income and Program Participation (SIPP). By using these two large, nationally representative datasets the power of estimates of observed effects can be improved over those found using smaller samples. Thus, this paper can obtain more precise estimates of the impact of income on morbidity for the working age population than has previously been possible in the literature.

3.2 Background on the Earned Income Tax Credit

The EITC program was first enacted in 1975 as a relatively small credit capped at $400 per family to offset payroll tax payments by families with children (Ventry 2001). Since then, Congress has enacted several expansions to the EITC program – with one of the most notable expansions occurring from 1993-1996 when the government introduced benefits for individuals without children and increased in the scope and generosity of the program for those with children. The expansions of the program from its initial size broadened the population claiming the EITC from 6.2 million families in 1975 to 24.6 million families in 2007. Similarly, the total value of annual benefits paid through the EITC program increased from 4.4 billion (2007 dollars) in 1975 to 48.5 billion dollars in 2007. (Tax Policy Center, 2009a).

A family's eligibility for federal EITC benefits is primarily dependent on the number of children in the family and the total labor earnings of all family members. In 2008, a single individual with two children and no labor earnings income receives no EITC benefits. Benefits are then phased in at a rate of 40 percent of labor earnings income for the first $12,060 of earnings – providing a maximum possible credit of $4,829. This maximum benefit is maintained until the family reaches $15,740 in labor earnings. For each dollar of labor earnings beyond $15,740 the EITC benefits are phased out at a rate of 21.06 percent of labor earnings. Once the family reaches $37,783 in labor earnings the benefits are completely phased out and the family is no longer eligible for any EITC benefits. The EITC benefits system is similar for families

Table 3.1: Maximum Federal EITC benefits and thresholds by year and number of children (in 2008 dollars)

	No children		One child		Two or more children	
	Earnings range for maximum benefits	Maximum federal benefits	Earnings range for maximum benefits	Maximum federal benefits	Earnings range for maximum benefits	Maximum federal benefits
1992			11,307 - 17,802	1,991	11,307 - 17,802	2,081
1993			11,371 - 17,901	2,104	11,371 - 17,901	2,217
1994	5,746 - 7,183	440	11,134 - 15,803	2,928	12,104 - 15,803	3,632
1995	5,752 - 7,197	440	8,641 - 15,838	2,938	12,121 - 15,838	4,363
1996	5,766 - 7,215	441	8,650 - 15,865	2,941	12,148 - 15,865	4,859
1997	5,805 - 7,263	444	8,694 - 15,957	2,956	12,225 - 15,957	4,890
1998	5,883 - 7,348	450	8,812 - 16,173	2,996	12,387 - 16,173	4,955
1999	5,854 - 7,327	448	8,787 - 16,101	2,988	12,328 - 16,101	4,931
2000	5,764 - 7,214	441	8,652 - 15,866	2,942	12,153 - 15,866	4,861
2001	5,789 - 7,236	443	8,683 - 15,919	2,953	12,186 - 15,919	4,874
2002	5,876 - 7,360	450	8,821 - 16,181	2,999	12,387 - 16,181	4,955
2003	5,842 - 7,305	447	8,768 - 16,073	2,982	12,304 - 16,073	4,922
2004	5,813 - 7,284	445	8,731 - 16,004	2,968	12,254 - 16,004	4,901
2005	5,757 - 7,202	440	8,636 - 15,849	2,936	12,132 - 15,849	4,853
2006	5,745 - 7,198	440	8,628 - 15,815	2,934	12,110 - 15,815	4,844
2007	5,805 - 7,269	444	8,712 - 15,981	2,962	12,243 - 15,981	4,897
2008	5,720 - 7,160	438	8,580 - 15,740	2,917	12,060 - 15,740	4,824

Source: Author's calculations based on Tax Policy Center (2009b)

with no children or one child but with lower maximum benefits and different thresholds for the phasing-in and phasing-out of benefits.[30] The structure of the EITC benefits is similar in earlier years as that for 2008, except that the benefits thresholds and rates of phasing in and out benefits vary by year. Table 3.1 illustrates the maximum benefits and thresholds for obtaining these benefits in each year since 1992.

In addition to the federal EITC benefits, a number of states supplement federal benefits with additional state-level EITC benefits. The first state to offer state-level supplemental EITC benefits was Rhode Island in 1986 and over the course of the 1990s and 2000s, supplemental state EITC programs expanded to a geographically

[30] Starting in 2009, the EITC program began offering benefits for families with three or more children above that provided for families with two children. Under the new policy, families with 3 or more children earned 45 percent of their first $12,590 in labor earnings in credits, an increase from the 40 percent of the first $12,590 in labor earnings provided to families with two children. Prior to this, all families with two or more children were treated identically regarding their EITC benefit eligibility.

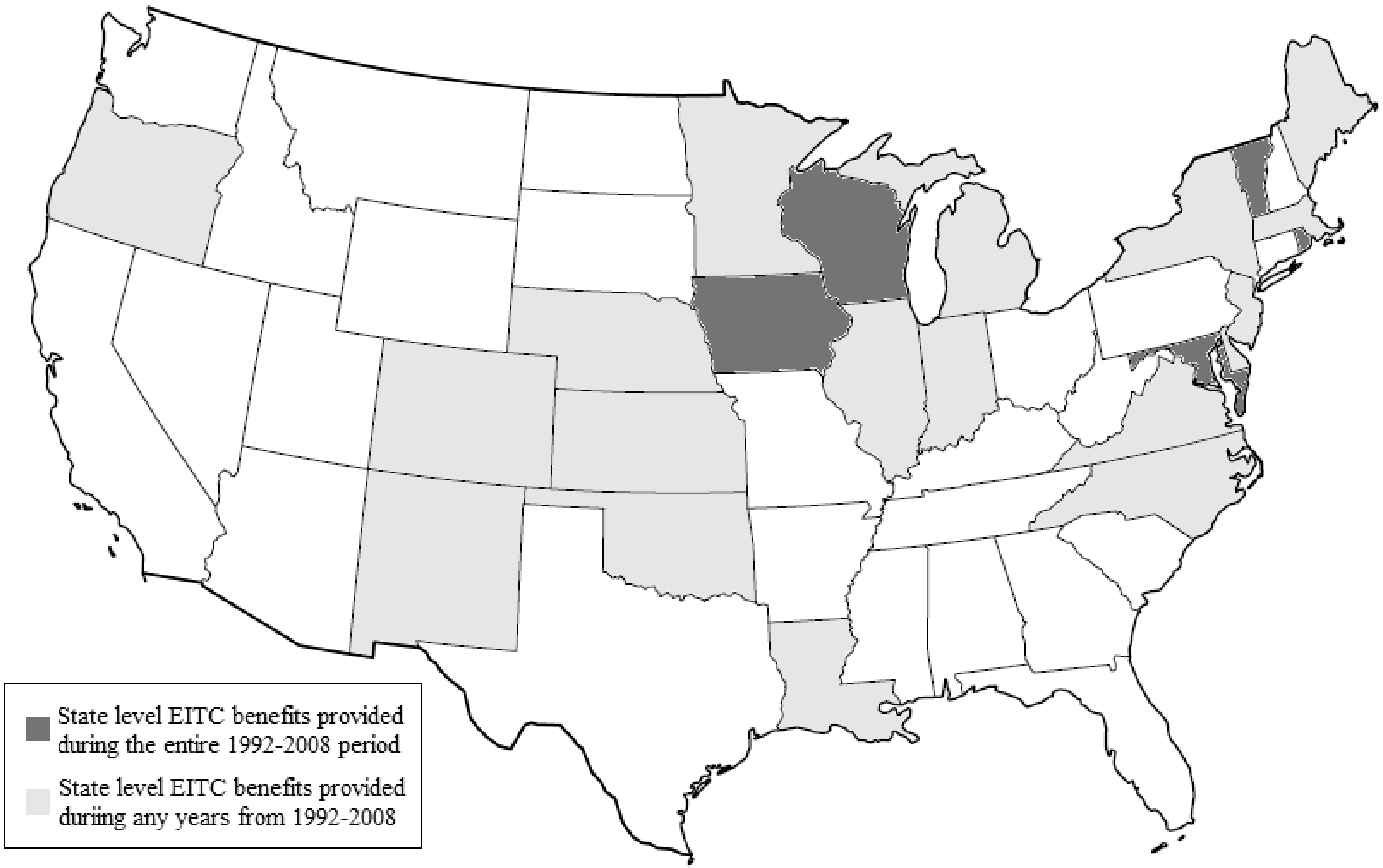
and politically diverse set of states. In 1992, the first year of this study, 5 states offered supplemental benefits and by 2008 23 states plus the District of Columbia had enacted state-level supplements to the federal EITC benefits (see Figure 3.1 for a map of states with state-level supplemental benefits). While there are differences in the refundability of state-level supplemental EITC benefits, these state-level benefits are typically calculated as a fixed percentage of a family's federal benefits. [31] In 2008, among states offering supplemental EITC benefits, the generosity varied widely from of 3.5 percent of federal benefits in Louisiana to 40 percent in the District of Columbia. For low income families this combination of state and federal benefits can be quite sizeable, representing between 7.65 percent and 56 percent of labor earnings for families with labor earnings below the disregard threshold.

3.3 Data

The primary data used in this paper comes from seven panels of SIPP data spanning the 14 years from 1992 through 2005.[32] The SIPP dataset is a nationally representative survey conducted by the Census Bureau. It follows at least 44,000 individuals for two to four year panels (starting in 1996 the SIPP expanded to approximately 110,000 individuals in each panel). Respondents are interviewed every three months on their income over the previous quarter and approximately once per year they are provided a topical question module asking about their health status and functional limitations. In addition to the SIPP data from 1992 through 2005, estimates are also calculated using data from the March CPS from 1996 through 2009. The March CPS interviews at least

[31] In addition to varying on the percentage of federal EITC benefits paid to recipients, states also differ on whether the credit is refundable. Given that many individuals eligible for the EITC pay little or no state income tax, refundable credits have the potential to be much more valuable to the target population.

[32] The SIPP does not ask about health status in questionnaires collected in 2006, 2007, or 2008, thus limiting the analysis to years prior to 2006. Additionally, no health questions were asked in 1995, which was the final year of the 1993 panel so no observations are recorded for this year.



Source: Author’s compilation from Leigh (1994), Feenberg (2007), and Tax Policy Center (2009c)

Figure 3.1: State level supplemental EITC benefits (1992-2008)

130,000 individuals each year (at least 200,000 each year since 2002), asking about their income over the previous calendar year and their current health status. However, unlike the SIPP, the March CPS does not include questions on functional limitations and it can only be used to consider the relationship between income and morbidity after 1996 when health status questions were added to the survey.[33]

Measuring Morbidity. In the summer or fall of each year respondents in the SIPP are asked "Would you say your health in general is excellent, very good, good, fair, or poor?" to obtain a measure of self-reported health status. The same question wording is also used in the March CPS to ask about health status. This question wording is the same or very similar to that used in numerous other surveys with self-reported health components, including the Health and Retirement Study and the Behavioral Risk Factor Surveillance System. Self-reported health status has widely been used in the income-health gradient literature (see, e.g., Case, Lubotsky and Paxson, 2002; Ettner, 1996; and Lindahl, 2005) and has been shown to be a good predictor of functional limitations (Idler and Kasl 1995; Idler, Russell, and Davis 2000), health care utilization (DeSalvo, et al. 2005) and future mortality (Wannamethee and Shaper, 1991, Idler and Kasl 1991; Idler, Russell, and Davis 2000, DeSalvo, et al. 2005).

Although self-reported health status is widely used to measure morbidity, it is a subjective measure of health which may vary based on the respondent's personal assessment of the scale. It is possible that two individuals with identical health could report different health ratings simply because of differences in their perceptions of good health. Therefore, to test the sensitivity of results to this measure of health status,

[33] One health-related variable that is available in the March CPS prior to 1996 is whether the respondent reports having a work-limiting disability. This question was asked in the March CPS beginning in 1980 and may be a valuable alternative measure to use in future research on the relationship between income and morbidity.

results are also calculated measuring health based on eight functional limitations which are asked in several years of the SIPP panel. These functional limitations are having difficulty seeing and reading newspaper print even with glasses or contact lenses, hearing normal conversation even when wearing a hearing aid, lifting and carrying 10 pounds, walking a quarter-mile, climbing a flight of 10 stairs, getting in and out of bed or a chair, doing light housework, and using an ordinary telephone. Each functional limitation is asked as part of a series of questions in the SIPP survey inquiring as to whether the respondent has any difficulty with the activity. They are both evaluated separately and using a single binomial scale where individuals are considered to have a functional limitation if they answer affirmatively to having one or more of the eight limitations.

Although the analysis of functional limitations is a valuable check of the results in this study, functional limitations are included only in the SIPP data and are not part of the CPS questionnaire. Therefore, for analyses of self-reported health status both the CPS and SIPP data are used, while analyses of functional limitations are limited to the SIPP dataset.

Measuring Income. Both the SIPP and March CPS contain extensive income questionnaires that are intended to capture most sources of cash income in the family. The batteries of income questions inquire about wages and salaries as well as income from interest, dividends, public transfers, and other sources of non-labor income. Similar to most income datasets, both the SIPP and the March CPS capture pre-tax income and do not ask about tax liabilities. Given that the majority of research on the income-health gradient focuses on this pre-tax income, this paper initially does the same. However, since the relationship between health and income is more likely to depend on income available to individuals for consumption which is more closely captured by post-tax income, it is also valuable to consider the relationship between post-tax income and morbidity. In order to explore post-tax income, taxes are imputed using NBER TAXSIM v9 based off of the income information provided by SIPP and CPS respondents. These imputed taxes are then added to each family's income to determine post-tax income for use in this analysis.[34]

Additionally, the resources available to any individual in a family depend both on the family's income and on the number of individuals in the family who share that income. Thus, throughout this paper income is adjusted for family-size by dividing by the square root of the number of individuals in the family.[35] To focus on low income individuals who will be affected by EITC benefit changes, the sample is restricted to individuals with pre-tax size-adjusted family income less than twice the federal poverty line for a single individual. This threshold roughly coincides with the maximum earnings families with 1 or 2 children can receive before EITC benefits are completely phased out. Thus, only individuals with pre-tax size-adjusted family income below $21,994 (2008 dollars) are included.

One limitation of using SIPP data in conjunction with EITC and tax information is that the interviews for each SIPP panel are staggered across months. This means that health questions are asked at different times in the calendar year for each respondent. Taxes and EITC benefits, however, are based off of calendar year incomes. To ensure that the income for each individual is considered for the same time-span prior to being asked their health status, taxes and EITC benefits are calculated using calendar years and are assumed to be paid or received equally across

[34] Since the EITC is based on the tax unit, which is generally more restrictive than a family as defined by the Census, a family is defined throughout this paper as the Census subfamily. While the traditional Census family includes all related individuals in the household, the subfamily separates each nuclear family into a separate subfamily which is closer to the expected tax-units.

[35] Dividing by the square-root of the family size is the most commonly used case of the economies of scale size-adjustments proposed by Buhmann et al. (1988) where Size-adjusted Family income = (total Family income) / (Family size)$^{\alpha}$, with α=1 implying no economies of scale and α=0 implying infinite economies of scale. Setting α=0.5 closely matches the adjustments for family size implied by the Census Bureau poverty thresholds (Ruggles 1990).

months in the year.[36] So, if individuals are asked about their health status in the beginning of October, the tax liabilities incorporated into their post-tax income will be 9/12 of the current year taxes (January through September) and 3/12 of the previous year taxes (October through December). This procedure allows for income to be analyzed over consistent 12-month periods prior to health questions being asked, regardless of whether that 12-month period aligns with a calendar year. The CPS data, in contrast, asks about current health and income for the previous calendar year each March. Thus the observation period for income in the CPS accurately aligns with a calendar tax year for both the calculation of EITC benefits and annual tax liabilities.

3.4 Empirical Strategy

To confirm that a positive relationship between income and health exists in the March CPS and SIPP data, the standard Ordered Logit regression:

$$H_{ist} = \alpha + \beta_1 Y_{ist} + \beta_2 X_{it} + \beta_3 Z_s + \varepsilon_{ist} \tag{3.1}$$

can be used, where H_{ist} is the individual's self-reported health status; Y_{ist} is the size-adjusted family income over the previous 12 months (previous calendar year in the CPS); X_{it} is a vector of demographic variables including age, race, ethnicity, gender, education, marital status, number of children in the family, residence in a MSA, and health-insurance status; and Z_s is a vector of state dummy variables to account for time-invariant differences between states that influence the health of its residents. The regression is performed as an Ordered Logit regression to reflect the ordinal rather than cardinal relationship between health statuses.

[36] While Individuals qualifying for EITC benefits may opt to receive them throughout the year as they pay payroll taxes rather than waiting until filing taxes in April of the following year, the take-up rate on this option is only around 0-2 percent over the past two decades (Holt , 2008). To the extent that individuals anticipate the refund but do not apply for the advance EITC benefits, however, this income change may be reflected in their spending decisions. Nevertheless, the delay in collecting EITC benefits may lead to an overstatement of the immediate increase in post-tax income that results from an increase in EITC benefit generosity.

While the existence of a relationship between income and morbidity can be observed using a standard Ordered Logitt regression, given the multiple pathways that can impact this relationship it would be misleading to interpret the effect as causal based on these results. Exploring causality instead requires finding an exogenous variation in income or health to separate the contributing effects. The changes in state and federal EITC benefits that have occurred since 1992 provide a natural experiment to explore the impacts of such an exogenous variation in income. Schmeiser (2009) exploited these variations to consider how an exogenous increase in income impacts obesity rates among low income individuals and this paper uses a similar strategy to consider the impact on morbidity of these individuals.

In order to identify the causal influence of changes in income on morbidity, each individual's family income is instrumented for using the maximum combined state and federal EITC benefits that the family could potentially receive. This maximum benefit level is a state level variable reflecting the credits a family in the state could receive if their labor earnings were such to maximize their benefits. It depends only on the state of residence and number of children in the family and not on the labor earnings of family members. Data on state and federal benefits were compiled based on Leigh (2004), Feenberg (2007), and the Tax Policy Center (2009b, 2009c).

The validity of using the maximum combined state and federal EITC benefits as an instrument requires that these benefits are correlated with family income but uncorrelated with the unobserved determinants of morbidity status. Since the health of low income individuals has not traditionally been a factor in setting EITC benefits, it is unlikely that the unobserved determinants of an individual's health will be correlated with the state-level EITC benefits in a given year. However, this assumption could be violated if other state spending that influences health outcomes vary along with EITC benefits. One of the most notable possibilities is if health insurance status is unobserved and Medicaid changes occurred in conjunction with shifts in EITC benefits. But this problem is mitigated since individuals are asked about their health insurance status. Health insurance status is therefore included in all regressions, making health insurance an observable rather than unobservable determinant of morbidity. Nevertheless, since it cannot be tested whether the exclusion restriction is satisfied, it remains possible that concurrent changes in other public spending programs could influence the results.

Although the maximum EITC benefits are unlikely to be correlated with the unobserved determinants of morbidity, these benefits should be correlated with the income of low income families. In addition to the large direct supplement to income from the increase in benefits, increases in the EITC benefits also have been found to increase labor force participation and subsequent labor earnings (Meyer, 2002).[37] As will be demonstrated below, given both the higher marginal income for hours worked and the labor supply effects from increasing EITC benefits, the generosity of state level benefits are a powerful predictor of family income for low income families.

While the EITC benefits available in a state are predictive of income for the low income population, this is not the case for high-income individuals who are not eligible for the benefits. Therefore, to focus on individuals influenced by EITC policies, only individuals with size-adjusted pre-tax income below twice the federal poverty line for a single individual are included in this study. Additionally, the sample is limited to working age individuals aged 22 through 62, who are most likely to be induced to enter or remain part of the labor market given an increase in their take-

[37] Snyder and Evans (2006) suggest that employment itself may have positive health benefits which could lead to an overestimate of the positive effects on health of higher income if not controlled for. Thus, regressions were analyzed both with and without hours worked as a control variable, as is discussed in more detail in the results.

home pay from the EITC benefits.[38] Even though these limitations prevent the generalization of results to the entire population, since the income health relationship is believed to be strongest at low incomes it is particularly valuable to understand the direction of the relationship among these individuals.

To estimate how an increase in income influences the health of low income individuals, the family's size-adjusted income is first estimated using the first stage equation:

$$Y_{ist} = \gamma + \delta_1 EITC_{ist} + \delta_3 X_{it} + \delta_4 Z_s + \mu_{ist} \tag{3.2}$$

where Y_{ist} is the size-adjusted family income over the previous 12 months (previous calendar year in the CPS); $EITC_{ist}$ is the maximum EITC benefits for which the individual could be eligible based on their number of children, state of residence, and year of observation; X_{it} and Z_s are the same vectors of demographic characteristics and state dummy variables described in Equation (3.1).

As discussed above, since the SIPP data is structured as a monthly dataset where individuals enter the sample in different months, the period of analysis does not generally represent calendar years. Potential EITC benefits, however, are based on calendar year income. This problem can be overcome using the same procedure used for distributing estimated taxes paid across months. Just as it was assumed that taxes paid are distributed evenly throughout the year it can be assumed that individuals eligible for the maximum credits would collect them evenly throughout the year. In the first stage equation, EITC benefits are weighted based on the number of months of each calendar year for which income is observed. Thus, if income is observed for October 1991 through September 1992, $EITC_{ist}$ will equal 3/12 of the maximum possible EITC benefits in 1991 plus 9/12 of the maximum EITC benefits in 1992. In

[38] Individuals age 22 through 24 are only eligible for EITC benefits if they have children, while individuals age 25 through 62 are eligible for some benefits even without any children in their families.

the annual March CPS survey, the maximum EITC benefits are considered for the previous calendar year which matches the period of observation for the individual's income.

Using the estimated income from the first stage, measures of individuals' health are then estimated using the second stage equation:

$$H_{ist} = \alpha + \beta_1 \hat{Y}_{ist} + \beta_2 X_{it} + \beta_3 Z_{st} + \varepsilon_{ist} \qquad (3.3)$$

where the individual's health, H_{ist}, is measured as self-reported health or the presence of functional limitations and $\hat{Y}_{ist}$ is the estimated income from the first stage equation. As was the case when estimating equation (3.1), when the dependent variable is self-reported health an Ordered Logit regression is used to reflect the ordinal rather than cardinal relationship between health statuses. When the dependent variable is the presence of functional limitations, a Logit regression is used to reflect the binary nature of functional limitations. If increases in income for low income individuals receiving EITC benefits improve health outcomes, then this will be reflected by a positive coefficient β_1. Similar to equations (3.1) and (3.2), vectors of demographic variables, X_{it}, and a vector of state dummy variables, Z_s, are included as control variables in the second stage regressions.

3.5 Results

The positive relationship between pre-tax income and self-reported health that has been well documented in the literature can be easily observed in both the SIPP and CPS data. Table 3.2 illustrates this relationship by providing the fraction of the working age population reporting each health status by decile of their pre-tax size-adjusted family income. In both datasets individuals in the bottom decile of income report being in poor health at over three times the rate of the total population. Similarly, the fraction of individuals in the bottom decile reporting being in excellent

Table 3.2: Frequency of self-reported health statuses by decile of pre-tax size-adjusted family income

Panel A: Survey of Income and Program Participation

Income Decile	Self-reported Health Status: Poor	Fair	Good	Very Good	Excellent
1	9.13	17.91	30.41	25.58	16.97
2	5.75	13.26	29.85	30.32	20.82
3	3.56	10.04	28.62	33.30	24.48
4	2.71	7.93	28.08	35.39	25.90
5	2.12	6.73	26.22	36.31	28.62
6	1.74	5.97	24.69	37.40	30.19
7	1.52	5.24	23.15	38.35	31.73
8	1.16	4.61	21.90	38.69	33.64
9	0.84	3.85	20.65	38.79	35.86
10	0.59	2.94	17.48	37.17	41.82
All	2.89	7.80	25.06	35.17	29.08

Panel B: March Current Population Survey

Income Decile	Self-reported Health Status: Poor	Fair	Good	Very Good	Excellent
1	10.25	16.51	29.29	24.76	19.20
2	6.97	13.03	29.28	28.82	21.89
3	4.08	9.71	28.33	32.78	25.10
4	2.89	7.79	26.94	34.41	27.97
5	2.23	6.54	25.43	35.61	30.19
6	1.73	5.72	24.27	36.26	32.02
7	1.41	4.99	22.66	37.07	33.87
8	1.15	4.44	21.93	36.76	35.71
9	1.00	3.82	19.65	36.86	38.67
10	0.77	3.23	16.66	35.22	44.11
All	3.25	7.58	24.44	33.85	30.87

Source: Author’s calculations based on SIPP and March CPS data files

health is less than 2/3 of that of the total population in both the SIPP and CPS data.

When using the standard Ordered Logit regression from equation (3.1) to regress self-reported health on pre-tax income for the entire working age population, including controls for demographic characteristics such as age, race, education, and marital status, this significant positive relationship is still observed (Columns 1 and 2

Table 3.3: Ordered Logit results regressing self-reported health on size-adjusted family income and demographic controls for working age individuals

	(1) SIPP All working age[1]	(2) CPS All working age[1]	(3) SIPP Low income[2]	(4) CPS Low income[2]
Pre-tax income ($1000s)	0.0087***	0.0052***	0.0160***	0.0138***
	(0.00013)	(0.00004)	(0.00124)	(0.00052)
Observations	307585	1325960	85397	356432

	(1) SIPP All working age[1]	(2) CPS All working age[1]	(3) SIPP Low income[2]	(4) CPS Low income[2]
Post-tax income ($1000s)	0.0154***	0.00876***	0.0191***	0.0091***
	(0.00022)	(0.00008)	(0.00140)	(0.00064)
Observations	307585	1325960	85397	356432

*significant at 10% level, **significant at 5% level, ***significant at 1% level
Source: Author's calculations based on SIPP and March CPS data
Note: Additional covariates in all regressions include gender, age, age-squared, race, ethnicity, education, year, state of residence, residence in an MSA, number of children in the family, marital status, and health insurance status.
[1]Working age population includes all individuals age 22-62
[2]Low income population includes all individuals of working age with size-adjusted pre-tax family income less than twice the federal poverty level for a single individual

of Panel A of Table 3.3). Additionally, the relationship between pre-tax income and self-reported health is even stronger when restricting the sample to low income working age individuals making less than twice the federal poverty line (Columns 3 and 4 of Panel A of Table 3.3).

If income is influencing health through its ability to allow for greater consumption activities, one should expect that the relationship is stronger when using post-tax income.[39] As can be seen in Panel B of Table 3.3, this is generally the case. Using the SIPP data for both all individuals and low income individuals, the

[39] Assuming income is influencing health through consumption, a valuable exercise would be to observe the changes in consumption occurring after an income shock and categorize their potential health effects. Previous research has used the Consumer Expenditure Survey to establish that consumption increases in response to tax cuts (Souleles 2002) and tax refunds (Souleles 1999), but to my knowledge no research has attempted to analyze the health effects of consumption shifts in response to tax policy changes.

coefficient for the relationship between income and health is greater using post-tax income than it was using pre-tax income. For the CPS data, this is true for all working age individuals, although not when the sample is restricted to only those with low incomes.

Additionally, similar to that seen using pre-tax income, the income-health relationship is stronger among low income individuals than it is for the population as a whole. Therefore, given the strength of the relationship at the lower tail of the income distribution, it is apparent that the low income population will be the primary focus going forward is particularly relevant to understanding the income-health gradient.

IV Regression: Pre-tax income and self-reported health. While there is a significant positive relationship between pre-tax income and self-reported health for low income individuals in the Ordered Logit regression, this does not provide insight into on the direction of causation for the reasons described above. To consider the extent to which health status is influenced by changes in pre-tax income, the generosity of state and federal Earned Income Tax Credit Benefits is used to instrument for income. As can be seen from the first-stage F-statistics in Table 3.3, the generosity of state and federal Earned Income Tax Credit Benefits meet the requirement of an F-statistic greater than 10 signifying a strong instrument. However, when using the IV approach, the significant positive relationship observed in the Ordered Logit regression largely disappears.

As can be seen in Column 1 of Table 3.4, when using the generosity of state and federal EITC benefits as an instrument for income in the SIPP data, higher pre-tax income has no significant effect on individuals' self-reported health status. The point estimates for the effect of income on morbidity are larger than those in the initial Ordered Logit regression, but because of the increase in standard errors from the IV approach the estimated effect is not significantly different from zero.

Table 3.4: Instrumental Variable results regressing self-reported health on pre-tax size-adjusted family income and demographic controls for low income working age individuals

	(1) SIPP	(2) SIPP	(3) CPS	(4) CPS
Pre-tax income ($1000s)	0.0592 (0.0671)	0.0394 (0.0723)	0.0757* (0.0442)	0.0519 (0.0750)
Employment Status	No	Yes	No	Yes
Observations	85397	85397	356432	356432
First stage F-statistic	28.42	26.66	49.12	19.68

*significant at 10% level, **significant at 5% level, ***significant at 1% level
Source: Author's calculations based on SIPP and March CPS data
Note: Additional covariates in all regressions include gender, age, age-squared, race, ethnicity, education, year, state of residence, residence in an MSA, number of children in the family, marital status, and health insurance status.

Given that a portion of the higher income from the EITC is due to an increase in hours worked, it is possible that the health effects of employment may be contributing to these results, since the regression in Column 1 of Table 3.4 does not control for employment status. However, previous research suggests that employment tends to improve individual's self-reported health and mortality outcomes (Gallo et al., 2000 and Snyder and Evans 2006). Thus, if changes in the EITC produce employment effects along with income effects, not controlling for the employment effects would likely overstate rather than understate the impact of higher income on health status.

When adding controls for whether the individual does not work, works part time, or works full time, the point estimate for the effect of income on health is lower than when these controls are excluded (Column 2 of Table 3.4). Thus, even with the controls for employment status, the observed effect of income on self-reported health is still not significantly different from zero and it does not appear that changes in individuals' employment statuses are a major factor for this result.

To validate the results from the SIPP data finding no evidence that changes in pre-tax income for low income individuals have an effect on morbidity, the previous

IV regression was replicated using the March CPS (Columns 3 and 4 of Table 3.4). When doing so, there was some support for a positive health effect of higher incomes, although these findings were relatively weak. The point estimates for the effect of changes in income on morbidity are greater than those found in the SIPP both with and without employment controls and when no employment control is included, the higher estimated effect of pre-tax income observed in the March CPS data is significantly positive at the 10% level. However, once including controls for the individual's employment status the estimated effect of pre-tax income on self-reported health declines and is no longer statistically significant. Therefore, while there is some evidence that pre-tax income positively influences self-reported health, the evidence supporting this theory is weak given the lack of any significant effects in the SIPP data and a positive effect only for one of the model specifications in the March CPS.

IV Regression: Post-tax income and self-reported health. Given that post-tax income more closely approximates the disposable income available to individuals for health-related consumption, would focusing on post-tax rather than pre-tax income influence the results? Although the first stage regression using the maximum potential EITC benefits to predict pre-tax income satisfied the standard requirement of an F-statistic greater than 10, the strength of the first stage relationship increases dramatically when using post-tax income instead. This is because the actual EITC benefits received are only observed when using post-tax income. This is in contrast to pre-tax income, which only captures the additional income from shifts in behavior after a change in EITC benefits and not the direct income from the benefits themselves. However, while focusing on post-tax income makes the maximum potential EITC benefits a stronger predictor of income, the results for how income impacts self-reported health are largely unchanged.

The results of the IV regression estimating the effect of changes in post-tax

Table 3.5: Instrumental Variable results regressing self-reported health on post-tax size-adjusted family income and demographic controls for low income working age individuals

	(1) SIPP	(2) SIPP	(3) CPS	(4) CPS
Post-tax income ($1000s)	0.0154	0.0097	0.0282*	0.0128
	(0.0174)	(0.0179)	(0.0164)	(0.0185)
Employment Status	No	Yes	No	Yes
Observations	85397	85397	356432	356432
First-stage F-statistic	540.42	571.43	521.58	459.40

*significant at 10% level, **significant at 5% level, ***significant at 1% level
Source: Author's calculations based on SIPP and March CPS data
Note: Additional covariates in all regressions include gender, age, age-squared, race, ethnicity, education, year, state of residence, residence in an MSA, number of children in the family, marital status, and health insurance status.

income on self-reported health are provided in Table 3.5. When using post-tax income, there is no change to the direction and significance of the key results although the point-estimates for the effect of income on self-reported health declines in both the SIPP and CPS data.

Similar to the findings using pre-tax income, the estimates derived from the SIPP data find that income changes resulting from shifts in the generosity of EITC benefits has no significant impact on health status. Using the March CPS data, the point estimates for the effect are once again larger than those in the SIPP and when no employment control is included the effect of post-tax income on self-reported health is positive and significant at the 10 percent level. But when adding the employment controls the estimated effect declines and is no longer significant. Therefore, mirroring the results using pre-tax income, while there is some evidence that post-tax income positively influences self-reported health, the evidence supporting a positive income effect on health is weak given the lack of any significant effects in the SIPP data and a positive effect only for one of the model specifications in the CPS.

IV Regression: measuring morbidity using functional limitations. While self-reported health status is commonly used to measure morbidity in survey data, an alternative and somewhat less subjective approach is to measure morbidity using self-reported functional limitations. Eight functional limitations which were included in each SIPP panel from 1992 through 2005 are considered in this paper. As was the case for self-reported health, when examining the prevalence of functional limitations among working age individuals by decile of the pre-tax size-adjusted family income distribution, each of the limitations are most prevalent among individuals in the bottom income decile (Table 3.6). The probability that a working age individual in the lowest decile of the income distribution reports at least one of these limitations is 2.2 times that of the general working age population.

Table 3.7 uses a Logit regression to estimate the relationship between pre-tax income and functional limitations among low income working age individuals. The outcome variable is having the specified functional limitation, so a negative coefficient signifies that increases in income reduce the probability that the individual reports the limitation and thus that higher income is associated with better health. For seven of the eight functional limitations and for the aggregated functional limitations variable, higher income is associated with lower rates of the limitation. This is true both using pre-tax (Panel A) and post-tax (Panel B) income. The one exception was having difficulty using a telephone, where the effect was reversed and higher income was associated with higher prevalence of the functional limitation.

To gain insight into the direction of these effects within the low income population, the same IV approach used previously is used here to re-estimates the effect of higher incomes on functional limitations. As was done when measuring morbidity using self-reported health, the effects are estimated using both pre-tax and post-tax income and with and without controls for employment status. The results using pre-tax income are provided in Panel A of Table 3.8 and those using post-tax

Table 3.6: Frequency of self-reported functional limitations by decile of pre-tax size-adjusted family income

Income Decile	reading newsprint	hearing conversation	lifting 10 pounds	climbing 10 stairs	walking 1/4 mile	using a telephone	getting out of bed or a chair	performing housework	any functional limitation
1	6.30	3.85	13.04	15.13	15.04	1.58	4.48	5.21	24.65
2	4.54	3.08	8.46	10.37	10.37	1.26	3.04	3.79	18.11
3	3.00	2.46	5.76	7.01	7.24	0.74	2.10	2.29	13.29
4	2.34	2.60	4.54	5.61	5.48	0.65	1.48	1.64	11.25
5	1.93	2.13	3.56	4.60	4.65	0.45	1.17	1.30	9.49
6	1.70	2.28	3.22	3.94	4.11	0.51	1.11	1.14	9.07
7	1.37	1.97	2.70	3.49	3.52	0.39	0.90	0.97	7.96
8	1.06	1.63	2.29	3.12	2.98	0.27	0.76	0.91	6.67
9	1.21	1.90	2.05	2.30	2.51	0.21	0.57	0.68	6.52
10	0.96	1.51	1.47	2.02	2.04	0.34	0.57	0.48	5.24
All	2.44	2.34	4.71	5.76	5.79	0.64	1.62	1.84	11.22

Source: Author's calculations based on the SIPP (1992-2005) data files

Table 3.7: Ordered Probit results regressing functional limitations on size-adjusted family income and demographic controls for low income working age individuals

	reading newsprint	hearing conversation	lifting 10 pounds	climbing 10 stairs	walking 1/4 mile	using a telephone	Getting out of bed or a chair	performing light housework	any functional limitation
Pre-tax income ($1000s)	-0.021***	-0.013***	-0.033***	-0.032***	-0.027***	0.020**	-0.030***	-0.020***	-0.023***
	(0.004)	(0.005)	(0.003)	(0.003)	(0.003)	(0.009)	(0.005)	(0.005)	(0.002)
Observations	58943	58943	58943	58943	58943	58943	58943	58943	58943

	reading newsprint	hearing conversation	lifting 10 pounds	climbing 10 stairs	walking 1/4 mile	using a telephone	Getting out of bed or a chair	performing light housework	any functional limitation
Pre-tax income ($1000s)	-0.028***	-0.017***	-0.043***	-0.040***	-0.037***	0.020**	-0.043***	-0.031***	-0.031***
	(0.005)	(0.006)	(0.004)	(0.003)	(0.003)	(0.010)	(0.005)	(0.005)	(0.002)
Observations	58943	58943	58943	58943	58943	58943	58943	58943	58943

*significant at the 10% level, **significant at the 5% level, ***significant at the 1% level

Additional covariates in all regressions include gender, age, age-squared, race, ethnicity, education, year, state of residence, residence in an MSA, number of children in the family, marital status, and health insurance status.

Source: Author's calculations based on the SIPP (1992-2005) data files

Table 3.8: Instrumental Variable results regressing self-reported health on size-adjusted family income and demographic controls for low income working age individuals

Panel A: Regression results using Pre-Tax Income

	reading newsprint	hearing conversation	lifting 10 pounds	climbing 10 stairs	walking 1/4 mile	using a telephone	getting out of bed/chair	performing light housework	any functional limitation
Pre-tax income ($1000s)	-1.328***	-2.687***	-0.308	-0.380	-0.434	0.381	-0.633	-0.229	-0.697***
	(0.488)	(0.550)	(0.354)	(0.343)	(0.337)	(1.036)	(0.585)	(0.560)	(0.261)
Employment Controls	No	No	No	No	No	No	No	No	No
Observations	58943	58943	58943	58943	58943	58943	58943	58943	58943
First-stage F-statistic	6.11	6.11	6.11	6.11	6.11	6.11	6.11	6.11	6.11

	reading newsprint	hearing conversation	lifting 10 pounds	climbing 10 stairs	walking 1/4 mile	using a telephone	getting out of bed/chair	performing light housework	any functional limitation
Pre-tax income ($1000s)	-1.421**	-3.235***	-0.027	-0.066	-0.112	0.919	-0.353	0.250	-0.579*
	(0.600)	(0.674)	(0.439)	(0.427)	(0.420)	(1.278)	(0.723)	(0.693)	(0.323)
Employment Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	58943	58943	58943	58943	58943	58943	58943	58943	58943
First-stage F-statistic	4.43	4.43	4.43	4.43	4.43	4.43	4.43	4.43	4.43

Table 3.8 (continued)

Panel B: Regression results using Post-Tax Income

	reading newsprint	hearing conversation	lifting 10 pounds	climbing 10 stairs	walking 1/4 mile	using a telephone	getting out of bed/chair	performing light housework	any functional limitation
Post-tax income ($1000s)	-0.185***	-0.375***	-0.043	-0.053	-0.060	0.053	-0.088	-0.032	-0.097***
	(0.068)	(0.077)	(0.049)	(0.048)	(0.047)	(0.144)	(0.082)	(0.078)	(0.036)
Employment Controls	No	No	No	No	No	No	No	No	No
Observations	58943	58943	58943	58943	58943	58943	58943	58943	58943
First-stage F-statistic	412.93	412.93	412.93	412.93	412.93	412.93	412.93	412.93	412.93

	reading newsprint	hearing conversation	lifting 10 pounds	climbing 10 stairs	walking 1/4 mile	using a telephone	getting out of bed/chair	performing light housework	any functional limitation
Post-tax income ($1000s)	-0.167**	-0.380***	-0.003	-0.008	-0.013	0.108	-0.042	0.029	-0.068*
	(0.070)	(0.079)	(0.052)	(0.050)	(0.049)	(0.150)	(0.085)	(0.081)	(0.038)
Employment Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	58943	58943	58943	58943	58943	58943	58943	58943	58943
First-stage F-statistic	433.18	433.18	433.18	433.18	433.18	433.18	433.18	433.18	433.18

*significant at the 10% level, **significant at the 5% level, ***significant at the 1% level

Additional covariates in all regressions include gender, age, age-squared, race, ethnicity, education, year state of residence, residence in an MSA, number of children in the family, marital status, and health insurance status.

Source: Author's calculations based on the SIPP (1992-2005) data files

income are provided in Panel B of Table 3.8.

Similar to the results found when measuring morbidity using self-reported health, in general higher income from increased EITC benefit generosity had no statistically significant effects on the prevalence of functional limitations. This was true for six of the eight functional limitations using both pre-tax and post-tax income. The two exceptions are whether the individual has difficulty reading newspaper print even when wearing glasses or contact lenses and whether the individual has difficulty hearing even when wearing a hearing aid. For each of these functional limitations, the increase in either pre-tax or post-tax income resulted in a significant reduction in the prevalence of the limitation. Additionally, the effect that higher income has on reducing these limitations is substantial enough that increases in income also significantly reduce the probability of reporting at least one of the eight limitations. One possible explanation for the more significant effect of increases in income on these particular limitations is the availability and relatively low cost of glasses, contact lenses, and hearing aids which are referenced in the questions and can easily mitigate the limitation. The vast majority of individuals with these limitations do not report being blind or completely deaf, but rather have limited vision or hearing. Thus, it is possible that their functional limitations are the result of having no glasses or hearing aids or an outdated prescription. But because each of these limitations can be easily corrected in the short run with the purchase of glasses or a hearing aid, the limitations can be corrected in the short-run more easily than the other functional limitations inquired about. Thus, these findings may be reflective of the direct health spending in the short-run that result from increases in income.

3.6 Conclusions

Numerous researchers have previously documented the positive relationship that exists between income and morbidity. Similarly, this paper observes this relationship over 13 years of data from the SIPP and March CPS even when controlling for the gender, age, race, ethnicity, and education of individuals. Using shifts in the generosity of state and federal EITC benefits to instrument for the income of low income individuals, it considers whether the relationship is derived from changes in income influencing morbidity rates. When doing so, there is only limited evidence that changes in income influence morbidity status.

Using both pre-tax and post-tax income, shifts in income among the low income population have only a statistically insignificant effect on self-reported health in the SIPP. Similarly, when measuring morbidity using functional limitations in the SIPP, changes in income resulting from EITC generosity have no significant effect on the prevalence of 6 of the 8 functional limitations. However, increases in income do appear to reduce the probability that an individual is unable to read newspaper print even when wearing glasses or contact lenses or hear normal conversation even when wearing a hearing aid. The positive significant effect of increases in income on these particular limitations would be consistent with individuals using additional income from their EITC benefits to remedy health ailments that are easily observable and relatively inexpensive to correct.

While the SIPP data finds no significant effect of shifts in income on self-reported health, a similar analysis in the March CPS provides limited support for a positive income effect. Using the March CPS data, the point estimates for the impact of higher pre-tax or post-tax income from increased EITC benefits are larger than those from the SIPP and are significantly positive when no employment controls are included. However, once employment controls are included the effect of higher income is no longer statistically significant so even in the March CPS the evidence for positive health effects from income are weak.

While the evidence in this study generally does not support the theory that higher incomes reduce morbidity rates in the short run, it is important to recognize that this does not rule out the possibility of long-term effects on health of an individual's income. Because both the SIPP and CPS follow individuals only for a relatively short period of time, it is not possible to observe the effects of higher income over several years or decades. Thus, over an extended period there may be long-term health effects from a permanent shift in income that exceed those observed just a year after the income shock occurs.

Despite this limitation, it is useful to understand the short-term impacts on health of from shifts in income. Many of the pathways through which one might envision income influencing health, including improved medical compliance, health-related behavioral changes, or reduced stress, would be expected to have an impact in the relatively short term even if these effects grow over time. However, it would nevertheless be valuable to further explore the long-run effects of income shocks and understand how these long-term effects compare to the short-run effects observed in this study.

REFERENCES

Backlund, Eric, Paul D. Sorlie and Norman J. Johnson. 1996. "The Shape of the Relationship between Income and Mortality in the United States: Evidence from the National Longitudinal Mortality Study." *Annals of Epidemiology* 6(1): 12-20.

Bound, John. 1989. “The Health and Earnings of Rejected Disability Insurance Applicants,” *American Economic Review* 79(3): 482–503.

Buhmann, Brigitte, Lee Rainwater, Guenther Schmaus and Timothy M. Smeeding. 1988. “Equivalence scales, well-being, inequality, and poverty: sensitivity estimates across ten countries using the Luxembourg Income Study (LIS) database.” *Review of Income and Wealth* 34(2): 115-142.

Case, Anne, Darren Lubotsky, and Christina Paxson. 2002. "Economic Status and Health in Childhood: The Origins of the Gradient." *American Economic Review* 92(5): 1308-1344.

Deaton, Angus S. 2002. "Policy Implications of the Gradient of Health and Wealth." *Health Affairs* 21(2): 13-30.

Deaton, Angus S. and Christina H. Paxson. 1998. "Aging and Inequality in Income and Health." *American Economic Review* 88(2): 248-253.

DeSalvo, Karen B., Vincent S. Fan, Mary B. McDonell and Stephen D. Fihn. 2005. “Predicting Mortality and Healthcare Utilization with a Single Question.” *Health Services Research* 40(4): 1234-1246.

Ettner, Susan L. 1996. "New Evidence on the Relationship between Income and Health." *Journal of Health Economics* 15(1): 67-85.

Feenberg, Daniel. 2007. “State EITC provisions 1977-2006.” Available online via: http://www.nber.org/~taxsim/eitc.html

Fuchs, Victor R. 1982. "Time Preferences and Health: An Exploratory Study," in

Fuchs, Victor R. (ed) *Economic Aspects of Health*. Chicago: University of Chicago Press.

Gallo, William T.; Elizabeth H. Bradley; Michele Siegel and Stanislav V. Kasl. 2000. "Health Effects of Involuntary Job Loss among Older Workers: Findings from the Health and Retirement Survey." *Journal of Gerontology Series B Psychological Sciences and Social Sciences* 55(3): S133-40.

Haveman, Robert; Barbara Wolfe; Lawrence Buron and Stephen C. Hill. 1995. "The Loss of Earnings Capability from Disability/Health Limitations: Toward a New Social Indicator." *Review of Income and Wealth* 41(3): 289-304.

Herd, Pamela; Robert F. Schoeni and James S. House. 2008. "Upstream Solutions: Does the Supplemental Security Income Program Reduce Disability in the Elderly?" *The Milbank Quarterly* 86(1): 5-45.

Holt, Stephen D. 2008. "Periodic Payment of the Earned Income Tax Credit." Brookings Institution Background Paper. Available online via: http://www.brookings.edu/~/media/Files/rc/papers/2008/0505_metroraise_supplement_holt/metroraise_supplement.pdf

Idler, Ellen L. and Stanislav V. Kasl. 1995. "Self-Ratings of Health: Do They Also Predict Change in Functional Ability?" *The Journals of Gerontology* 50(6): S344-353.

Idler, Ellen L.; Louise B. Russell and Diane Davis. 2000. "Survival, Functional Limitations, and Self-rated Health in the NHANES I Epidemiologic Follow-up Study, 1992." *American Journal of Epidemiology* 152(9): 874-883.

Leigh, Andrew. 2004. "*Who Benefits from the Earned Income Tax Credit? Incidence Among Recipients, Coworkers and Firms*." Discussion Paper 494, Center for Economic Policy Research, Australian National University.

Lindahl, Mikael. 2005. "Estimating the Effect of Income on Health and Mortality

Using Lottery Prizes as an Exogenous Source of Variation in Income." *Journal of Human Resources* 40(1): 144-68.

McDonough, Peggy, Greg J. Duncan, David Williams, and James House. 1997. "Income Dynamics and Adult Mortality in the U.S., 1972-1989." *American Journal of Public Health* 87(9): 1476-1483.

Norris, Jean C., Mark J. van der Laan, Sylvia Lane, James N. Anderson, and Gladys Block. 2003. "Nonlinearity in Demographics and Behavioral Determinants of Morbidity." *Health Services Research* 38(6):1791-1818.

Meyer, Bruce. 2002. "Labor Supply at the Extensive and Intensive Margins: The EITC, Welfare, and Hours worked." *American Economic Review* 92(2): 373-379.

Ruggles, Patricia. 1990. "Drawing the Line: Alternative Poverty Measures and Their Implication for Public Policy." Washington, DC: Urban Institute Press.

Ruhm, Christopher J. 2000. "Are Recessions Good for Your Health?" *Quarterly Journal of Economics* 115(2): 618-649.

Schmeiser, Maximilian D. 2009. "Expanding Waistlines and Wallets: The Impact of Family Income on the BMI of Women and Men Eligible for the Earned Income Tax Credit." *Health Economics* 18(11): 1277-1294.

Smith, James P. 1999. "Healthy Bodies and Thick Wallets: The Dual Relation between Health and Economic Status." *Journal of Economic Perspectives* 13(2): 145-166.

Smith, James P. 2004. "Unraveling the SES-Health Connection," *Population and Development Review* 30, supplement Aging Health and Public Policy: 108-132.

Snyder, Stephen E. and William N. Evans. 2006. "The Effect of Income on Mortality: Evidence from the Social Security Notch." *Review of Economics & Statistics*

88(3): 482-495.

Souleles, Nicholas S. 1999. “The Response of Household Consumption to Income Tax Refunds.” *American Economic Review* 89(4): 947-958.

Souleles, Nicholas S. 2002. “Consumer Response to the Reagan Tax Cuts.” *Journal of Public Economics* 85(1): 99-120.

Tax Policy Center. 2009a. “Historical EITC Recipients” Available online via: http://www.taxpolicycenter.org/taxfacts/displayafact.cfm?DocID=37&Topic2id=40&Topic3id=42

Tax Policy Center. 2009b. “Historical EITC Parameters” Available online via: < http://www.taxpolicycenter.org/taxfacts/displayafact.cfm?DocID=36&Topic2id=40&Topic3id=42

Tax Policy Center. 2009c. “State EITC Based on the Federal EITC” Available online via: http://www.taxpolicycenter.org/taxfacts/displayafact.cfm?Docid=293

Ventry Jr., Dennis J. 2001. “The Collision of Tax and Welfare Politics: The Political History of the Earned Income Tax Credit: 1969-99.” *National Tax Journal* 53(4): 983-1026

Wannamethee, Goya and A.G. Shaper. 1991. “Self-Assessment of Health Status and Mortality in Middle-Aged British Men.” *International Journal of Epidemiology* 20(1): 239-245

Zimmer, Zachary and James S House. 2003. "Education, Income, and Functional Limitation Transitions among American Adults: Contrasting Onset and Progression." *International Journal of Epidemiology* 32(6): 1089-1097.